IRE Transactions

on INFORMATION THEORY

Volume IT-3 DECEMBER, 1957 Number 4

In This Issue

Frontispiece	page 212
Editorial	page 213
Contributions	page 214
Correspondence	page 256
Contributors	page 258
Annual Index 1957	Follows page 258

For complete Table of Contents, see page 211.

PUBLISHED BY THE
Professional Group on Information Theory

IRE Professional Group on Information Theory

The Professional Group on Information Theory is an organization, within the framework of the IRE, of members with principal professional interest in Information Theory. All members of the IRE are eligible for membership in the Group and will receive all Group publications upon payment of an annual fee of $3.00.

IRE TRANSACTIONS® ON INFORMATION THEORY is published by the IRE for the Professional Group on Information Theory, at 1 East 79th Street, New York 21, N. Y. Responsibility for contents rests upon the authors and not upon the IRE, the Group, or its members. Price per copy: IRE-PGIT members, $1.55; IRE members, $2.35; nonmembers, $4.65.

IRE Transactions on Information Theory

Published Quarterly by the Professional Group on Information Theory

Volume IT-3 | December, 1957 | Number 4

TABLE OF CONTENTS

PAGE

Frontispiece — *W. B. Davenport, Jr.* 212

Editorial — *W. B. Davenport, Jr.* 213

Sputnik, Et Cetera

Contributions

A Theory of Multilevel Information Channel with Gaussian Noise — *Satosi Watanabe* 214

A Generalization of a Method for the Solution of the Integral Equation Arising in Optimization of Time-Varying Linear Systems with Nonstationary Inputs — *Marvin Shinbrot* 220

On the Mean Square Noise Power of an Optimum Linear Discrete Filter Operating on Polynomial Plus White Noise Input — *Marvin Blum* 225

The Distribution of the Number of Crossings of a Gaussian Stochastic Process — *Carl W. Helstrom* 232

An Analysis of Coherent Integration and Its Application to Signal Detection — *K. S. Miller and R. I. Bernstein* 237

The Sequential Detection of a Sine-Wave Carrier of Arbitrary Duty Ratio in Gaussian Noise — *H. Blasbalg* 248

Correction to "Optimum Detection of Random Signals in Noise, with Application to Scatter-Multipath Communication, I" — *Robert Price* 256

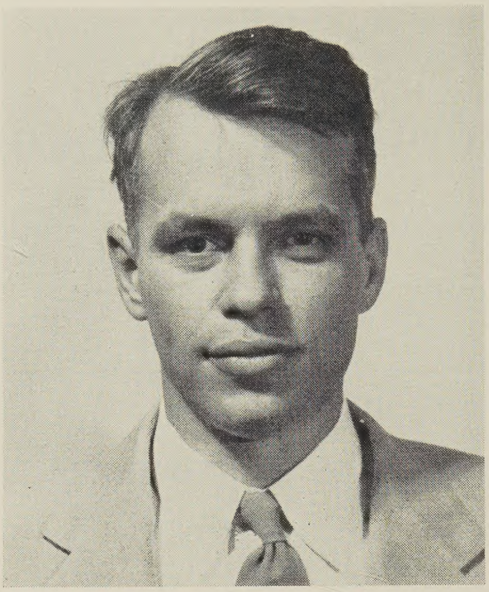
Correction to "On the Detection of Stochastic Signals in Additive Normal Noise—Part I" — *David Middleton* 256

Correspondence

The Sample Space Trajectory of Time-Shifted Signal Vectors — *Herbert Sherman* 257

Contributors 258

Annual Index 1957 — Follows Page 258



W. B. Davenport, Jr.

Wilbur B. Davenport, Jr., (SM '54) was born on July 27, 1920, in Philadelphia, Pa. He received the B.E.E. degree from the Alabama Polytechnic Institute, Auburn, Ala., in 1941, and the degrees of S.M. in E.E. and Sc.D. from the Massachusetts Institute of Technology in 1943 and 1950, respectively. From August, 1943, until August, 1946, he was on active duty in the United States Naval Reserve as a Fire-Control Radar Officer.

From 1946 to 1949 Dr. Davenport was an instructor in the Department of Electrical Engineering at M.I.T. He was promoted to assistant professor in that department in 1949, a position he held until 1953. In 1951 he also became Group Leader of the Communications Techniques Group at the M.I.T. Lincoln Laboratory. He was made Associate Head of the Communications and Components Division of Lincoln Laboratory in 1955, and was promoted to Head of that division in February, 1957.

Dr. Davenport has worked on the statistical properties of speech waveforms, on the effects of signals and noise in limiters, and on applications of the statistical theory of communications. He is co-author, with Dr. William L. Root, of the book "Random Signals and Noise."

He is a member of the Acoustical Society of America, Sigma Xi, Eta Kappa Nu, Tau Beta Pi, Phi Kappa Phi, Spiked Shoe, and the Armed Forces Communications and Electronics Association. Dr. Davenport has been a member of the PGIT Administrative Committee since 1954, Vice-Chairman in 1956, and Chairman in 1957.

Sputnik, Et Cetera

W. B. DAVENPORT, JR.

It is becoming a tradition of the IRE Professional Group on Information Theory to have its Chairman contribute an editorial to the December issue of these TRANSACTIONS. In keeping with this tradition, I would like to add a few words of my own to the daily growing clamor occasioned by the satellite Sputnik.

Ever since the first Russian satellite was launched into its orbit on October 4, 1957, a torrent of words has been inundating the daily press bearing charges and countercharges as to why the U.S. lags the U.S.S.R. in the development of missiles and satellites. The main topics of this discussion seem to be: 1) the importance of basic research, 2) the impact of the present and past Federal fiscal policies on military research and development programs, 3) the effect of interservice rivalries on ballistic missile development and procurement, and 4) the relative merits of the U.S. and the U.S.S.R. programs for the education of scientists and engineers. While all of these topics are important, I shall confine my remarks to the first—at the risk of ignoring the old Italian proverb, "A missionary makes few converts amongst true believers."

A cursory review of the statements of the past several months would give one the feeling that, of course, we are all in favor of basic research. Certainly the development of atomic and nuclear weaponry resulting from the early studies of the atomic nucleus, the development of the transistor resulting from basic studies of the solid state, and the practical developments resulting from the theories of Shannon and Wiener are all well-known stories and should provide incontrovertible evidence for the necessity of continually supporting free and untrammeled basic research. Unfortunately, the facts are otherwise. Basic research, or the gathering of new fundamental knowledge, is all too often treated as something to be dispensed with in times of crisis, not only by responsible public officials but also by many scientists and engineers themselves. I feel that this is a fundamental error. Sooner or later we are going to have to face the fact that crises will always be with us.

Regardless of what we do now to meet the crisis of the present, we must also maintain an active program of basic research in all the sciences now or we will be unable to meet the crises of the future. This must be kept in mind throughout our present necessary, and costly, rush to push ballistic missiles into production.

Therefore, may I add one small voice to the plea for reason. Basic research must never be abandoned in the interest of expediency, and in fact should be emphasized—in direct proportion to the increased effort applied to the opportune developments. This holds as much in our own field of Information Theory as in the other disciplines. Hence, it is my earnest hope that we shall effect an appropriate balance between basic advances and practical applications, as previous contributors to this space have entreated, and that we will not be unduly influenced by the force of temporal circumstance, expedience, or Sputniks.

A Theory of Multilevel Information Channel with Gaussian Noise*

SATOSI WATANABE†

Summary—The interval between the "0" level and "1" level of a binary channel with Gaussian noise is subdivided to provide $n = 2^g$ levels per symbol. The channel capacity is computed as a function of g and of the signal-noise ratio D of the original binary channel. For a sufficiently large, fixed value of D, if we increase g indefinitely, the channel capacity approaches the logarithm of D, as can be expected from continuous channel. For a fixed value of g, if we increase D indefinitely, the channel capacity approaches g. For a given value of D, there is a certain value of g, beyond which the channel capacity does not appreciably increase any longer by increasing g. The problem is first solved by a simplifying model, and then the error introduced by this simplification is estimated.

"Ring" Model

The original problem sketched in the Summary pertains, in brief, to the question as to how far we can expect to be able to improve the capacity of a given binary channel by splitting the level interval into more than one subinterval, when the binary channel has a (symmetric) probability of error p; *i.e.*, when its capacity is given by

$$C = 1 + p \log p + (1 - p) \log (1 - p), \tag{1}$$

where log stands for logarithm to the base 2. Thus, the problem may be characterized as that of an amplitude-modulated information channel. In order to obtain a qualitative insight into the general nature of such a channel, a simple Gaussian type of noise will be assumed.

Before coping with this problem, we shall introduce another one, which by itself is a meaningful one based on a set of well-defined assumptions, and whose solution is relatively easy, although it may not directly correspond to a realistic technological problem. We shall see in a later section how we can derive conclusions pertinent to the original problem from this "model," and we shall also evaluate the error committed by substituting this model for the original problem.

The problem we attack first is as follows: each symbol of the code is represented by one of the n available points separated by arcs of equal length $l = L/n$ on a circle of circumference L. We consider only those n's that satisfy $n = 2^g$ with g positive integers. The basic assumption we make is that the error distribution of the physical quantity implementing this channel, whose domain of variation here is the circle, is of the Gaussian type with standard deviation σ. Hereafter, we shall measure any arc of the circle by the unit of σ. Thus, the total circumference will be given by $G = L/\sigma$, and the interval by $d = l/\sigma = L/(n\sigma) = G/n$.

* Manuscript received by the PGIT, February 13, 1957.
† IBM Research Lab., Poughkeepsie, N. Y.

If any one of the n possibilities is transmitted; *i.e.*, if any one (call this one P_0) of the n points on the circle is sent, the corresponding point at the receiving end will be distributed with the probability distribution:

$$f(x) = (1/\sqrt{2\pi}) \exp(-x^2/2) \tag{2}$$

on the circle about the point P_0. It will be natural to assume that if the received point x falls on the interval: $rd - (1/2)d < x < rd + (1/2)d$ with r an integer, the receiver will recognize the signal as a point P which is rd apart from the original point P_0; *i.e.*, the rth neighbor of P_0. We shall identify the relative position of P by the integer r which is limited by: $-(1/2)n < r \leq (1/2)n$.

Then the probability of the transmitted signal P_0 being registered by the receiver as its rth neighbor will be given by

$$p(r) = \sum_{k=-\infty}^{\infty} \int_{rd-(d/2)+kG}^{rd+(d/2)+kG} f(x)\, dx, \tag{3}$$

where the integrand is given by (2). For instance, the term $k = 1$ in (3) represents the contribution from the error going, in one direction, one entire circumference plus rd. We have obviously

$$p(r) = p(-r); \quad \sum_r p(r) = 1, \quad -(n/2) < r \leq (n/2). \tag{4}$$

If we denote by $p_i(j)$ the probability of the transmitted signal i being received as signal j, and by w_i the weight with which signal i is used by the transmitter, then the theoretical channel capacity is given by[1]

$$C = \sum_j \sum_i w_i p_i(j) [\log p_i(j) - \log \sum_k w_k p_k(j)]. \tag{5}$$

In the following, we shall assume that each signal is used by the transmitter with equal weight; *i.e.*, $w_i = 1/n = 1/2^g$. Then (5) becomes

$$C = \log n + (1/n) \sum_j \sum_i p_i(j) \log p_i(j) - (1/n) \sum_j P(j) \log P(j), \tag{6}$$

with

$$P(j) = \sum_i p_i(j). \tag{7}$$

In our case, we have obviously $P(j) = 1$, hence the channel capacity is given by

[1] This is the channel capacity when the w's are fixed. The maximum of this quantity obtained by varying the w's is what is usually called channel capacity. See, for example, C. E. Shannon and W. Weaver, "The Mathematical Theory of Communication," University of Illinois Press, Urbana, Ill., p. 38; 1949.

$$C = g + \sum_r p(r) \log p(r), \qquad -(n/2) < r \leq (n/2). \quad (8)$$

It is easy to see that for extremely high noise and extremely low noise, (8) with (3) gives

$$\lim_{\sigma \to \infty} C = g - \log n = 0; \quad \text{and} \quad \lim_{\sigma \to 0} C = g. \quad (9)$$

Furthermore, we can expect, and we shall indeed demonstrate later, that for a given value of G and in the limit $g \to \infty$, we get a channel capacity per symbol equal to the logarithm of the signal-noise ratio G, since the nature of the problem approaches a continuous channel and the order of magnitude of the maximum frequency involved here is given by[2]

$$\nu_{\max} \approx 1/(\text{duration of a symbol}). \quad (10)$$

Approximation for Large G (Low Noise)

Without precipitating to the limiting case $\sigma \to 0$, $G \to \infty$, let us examine the cases where G is so large that the probability of error larger than the order of the circumference is negligible. Usual applications of the theory actually lie in this domain.

Under this assumption, all the errors propagating more than one circumference in one direction can be neglected. There are then only two possible paths to reach point r from P_0, namely one propagating clockwise and the other propagating counterclockwise from P_0 to r, neither being longer than the circumference G. Furthermore, unless r is in the neighborhood of $r = n/2$ (which is the point diametrically opposite to P_0), one of the paths is appreciably larger than the other. Due to the nature of the Gaussian distribution, the contribution to $p(r)$ from the shorter path will then be considerably larger than the contribution from the longer path. Thus we need consider only the shorter path for those points. In addition to this, if G is large enough, it is easy to see that $p(r)$ for r in the vicinity of $r = n/2$ is so small that its contribution to C in (8) is negligible compared with the contributions from the r's which are closer to the original point P_0. Writing $r = r_0$ $(n/2 < r_0 < 0)$ for the point which marks the limit of the "vicinity" of the point $r = n/2$, one may thus write

$$C = g + p(0) \log p(0) + 2 \sum_{r=1}^{r_0} p(r) \log p(r), \quad (11)$$

with

$$p(r) = p(-r) = \int_{rd-(d/2)}^{rd+(d/2)}, \quad (12)$$

where the omitted integrand is given by (2). Since the contribution $p(r) \log p(r)$ from r beyond $r = r_0$, whether or not $(n/2) > r$, is very small, one can write as well

$$C = g + \sum_{r=-\infty}^{\infty} p(r) \log p(r); \qquad \left(\sum_{r=-\infty}^{\infty} p(r) = 1\right), \quad (13)$$

with $p(r)$ given in (12). A careful estimation of errors shows that (13) is perfectly reliable for $G \geq 10$.

Approximation for Large g (Large Multiplicity), and Qualitative Description of the Actual Situation

We shall introduce first an approximation which is good for values of $d(= G/n = G/2^g)$ which are smaller than unity, with the restriction, however, that $G > G_0$, where G_0 is the value of G beyond which (13) becomes reliable; *i.e.*, G_0 is somewhere in the neighborhood of 10, say. For a given noise level, *i.e.*, for a given value of G, the approximation, therefore, will be good for large g. But for a given value of g, the approximation will become good for small G, within the limit $G > G_0$.

If $d \ll 1$, we can write (12) as

$$p(r) = (d/\sqrt{2\pi}) \exp[-(rd)^2/2]. \quad (14)$$

The summation in (13) can then be evaluated as follows:

$$\begin{aligned}\sum_{r=-\infty}^{+\infty} p(r) \log p(r) &= \log (d/\sqrt{2\pi}) \int_{-\infty}^{\infty} f(x)\,dx \\ &\quad - (1/2) \log e \int_{-\infty}^{\infty} x^2 f(x)\,dx \\ &= \log d - (1/2) \log (2\pi e) \\ &= \log G - g - (1/2) \log (2\pi e). \quad (15)\end{aligned}$$

Putting (15) in (13) we obtain

$$C_\infty = \log G - (1/2) \log (2\pi e). \quad (16)$$

Then the correct expression of the channel capacity, (8), or where applicable, (13) can be written as

$$C_g = C_\infty + A, \quad (17)$$

with

$$A = \sum p \log p - [\log d - (1/2) \log (2\pi e)]. \quad (18)$$

For a given value of G, the larger the value of g, the smaller the correction represented by A.

The condition

$$d \ll 1 \quad (19)$$

which is the premise for the derivation of (15), means[3]

$$G \ll 2^g \quad \text{or} \quad \log G \;\mathrm{E}\; g. \quad (20)$$

Actual calculation shows that A is negative, and $|A|$ is less than 3 per cent of the absolute value of $\sum p \log p = C_g - g$ at $d = 1$. Thus, already for $g = \log G$, the channel capacity C_g of (17) can very well be approximated by C_∞ of (16). For instance, for $G = 10$, C_g can tolerably be replaced by C_∞ for g equal to or larger than 3.

For very large values of G, the constant term in (16) can be neglected, and C_∞ becomes, as anticipated, the logarithm of the snr: $G = L/\sigma$.

[2] This matter will be discussed more fully in a separate paper.

[3] The symbol E is used in the sense that $x \;\mathrm{E}\; y$ means $2^{y-x} \gg 1 (x > 0, y > 0)$.

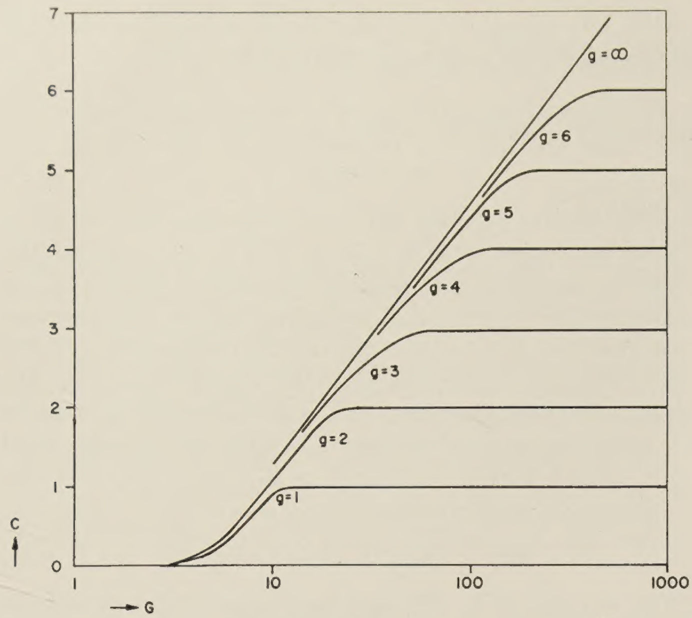


Fig. 1—Channel capacity C of the "ring" model vs the signal-noise ratio G for various g.

For a qualitative understanding of the situation, we should note the following facts. Considering C_g as a function of G for a given g, we can use C_∞ in lieu of C_g as long as $G_0 < G < 2^g$. As G increases beyond 2^g, C_g will gradually branch off the curve of C_∞ and level off toward the limiting value which is g. [See (9)]. For G smaller than G_0, C_g will approach zero as G decreases. [See (9)]. Fig. 1 shows exact curves of C_g for various g (including $g = \infty$). They are obtained using (3) and (8), or, where permissible, corresponding approximative formulas previously introduced.

In a still coarser approximation, we can describe the situation as follows: the exact curve of C_g for a given g lies very close to the curve C_∞ so long as C_∞ is less than g. Then the exact curve rather abruptly levels off at the limiting value g. A type of question which arises in practice pertains to how far we can improve the channel capacity by increasing g when the noise level G is given. A zeroth approximation answer to this question is as follows. For the given value of G, we calculate C_∞, and determine g_0 by

$$g_0 - 1 < C_\infty < g_0. \tag{21}$$

Then this g_0 will give the smallest value of g which gives a channel capacity very close to the maximum capacity, C. Any larger value of g will not appreciably improve the channel capacity. For instance, for G in the domain $17 < G < 33$, C_∞ lies between 2 and 3. Therefore, g larger than 3 will not improve the capacity to an appreciable extent in this domain.

For a little more precise discussion, we can search the smallest g which satisfies

$$0.9 < C_g/C_\infty \tag{22}$$

for the given value of G. Then any value larger than this g will give only an improvement within 10 per cent of C_∞.

Straight Multilevel Channel and Ring Model

The general idea underlying the proposed comparison of the ring model mentioned previously to the original problem of a straight multilevel channel, is that we cut the ring at a point and stretch it straight and compare the two end points on this strip to the "0" level and the "1" level of the original binary channel. Thus, we divide the circle G in $n = 2^g$ $(g > 1)$ equal intervals and cut it at one point and denote the distance between the two end points by D. The formulas obtained in the preceding sections were written in terms of G. It is more convenient hereafter to write them in terms of D which is obviously $G - d$. Therefore we should substitute for G

$$G = D2^g/(2^g - 1). \tag{23}$$

The ratio of D to G of course tends to unity as we increase g.

The general expression (17) of channel capacity in the ring model is now rewritten as

$$C_g = \log D - (1/2) \log (2\pi e) + A + B, \tag{24}$$

where A (< 0) is given by (18), and B is given by

$$B = g - \log (2^g - 1) > 0. \tag{25}$$

For very large g, we have

$$C_\infty = \log D - (1/2) \log (2\pi e). \tag{26}$$

This C_∞ and C_∞ of (16) becomes identical when g is really infinitely large. Otherwise, the difference is given by B.

Now, let us go back to the original problem. Suppose we are given a binary channel, in which a certain value q_0 and value 0 of a physical quantity q, *e.g.*, voltage, power, etc., are supposed to correspond respectively to the "1" and the "0" of the binary system. It is assumed that this physical quantity suffers during transmission fluctuation of the Gaussian type with a *constant* standard deviation σ (in the same unit as q) about the originally transmitted signal strength of q. If the physical quantity q itself does not show a Gaussian fluctuation, we should adopt a suitable function of q, for which the fluctuation is approximately Gaussian. Let us measure the "1" level; *i.e.*, the value q_0 of q by unit σ. Then what was called D previously corresponds to

$$D = q_0/\sigma. \tag{27}$$

However, there is obviously a difference between the ring model and the original straight multilevel problem, namely, in the former model errors beyond one end "overflow" to the other end. In the original problem, it will be natural to assume that any received signal beyond an end-point will be registered as belonging to this end-point. This discrepancy will become very quickly negligible as D increases. In fact, the relative number of levels participating in the "overflow" becomes smaller as D increases. For a given value of D, if we increase g, there will be more and more levels participating in the "over-

flow," but at the same time the number of levels which do not participate in the "overflow" will also increase proportionally. Therefore, the relative error will not appreciably depend on g (beyond a certain value of g) when D is once given.

Let us compare first the straight binary channel and the corresponding circular binary channel. In the latter, any received value beyond (1/2) D will be registered as "1", and any value below (1/2) D will be registered as "0". Thus, the probability of error is

$$p(1) = \int_{D/2}^{\infty} \equiv a. \tag{28}$$

Now in the ring model (with $g = 1$) with circumference G with two points $D = (1/2)G$ apart, we have (for $D > 1$).

$$p(1) = \sum_{k=-\infty}^{+\infty} \int_{kG-D/2}^{kG+D/2} \doteq 2\int_{D/2}^{3D/2} \doteq 2\int_{D/2}^{\infty} = 2a. \tag{29}$$

Thus, the capacities calculated with (28) and with (29), with the same D, are different. Therefore, when a straight binary channel with a certain probability of error p is given, it may be reasonable in this ring approximation to determine D by equating this p to $2a$ instead of a. However, on account of the nature of the error function, the value of D obtained by setting $p = a$ and the one obtained by $p = 2a$ do not differ appreciably. For instance, when the latter choice gives $D = 10$, the former choice gives $D = 9.6$. A difference in D of this order of magnitude does not result in an appreciable difference in its channel capacity. For larger values of D, the difference is entirely negligible.

Fig. 2 gives $2a$ as a function D, which will help to evaluate the appropriate value of D to be used in the multilevel problem. Fig. 3 shows C_g as given by (24) and C_∞ as given by (26), both as functions of D.[4]

Errors due to "Overflow"

The error committed by substitution of the ring model for the original "straight" channel is due to the overflows. One way to subdue this error is, as suggested in the preceding section, to use a modified value of D for the ring model; *i.e.*, to determine D by setting $p = p(1)$ in (29) instead of in (28). A more consistent way to attack this problem is to use the same D for the ring model and the original problem and to evaluate the difference in the resulting channel capacity.

For smaller values of g, direct calculation of this difference can easily be done. For instance, in the binary channel ($g = 1$), we can obtain the capacity C' of the straight channel by putting (28) in (1) and the capacity C of the ring model by putting (29) in (1).

[4] When G is used as abscissa as in Fig. 1, a curve with larger g does not come below a curve with smaller g. This situation is no longer a general rule when D is used as abscissa as in Fig. 3. It may, however, be expected that this situation will be reestablished even with D as abscissa if the overflowing errors are stopped and the w's in (5) are suitably chosen.

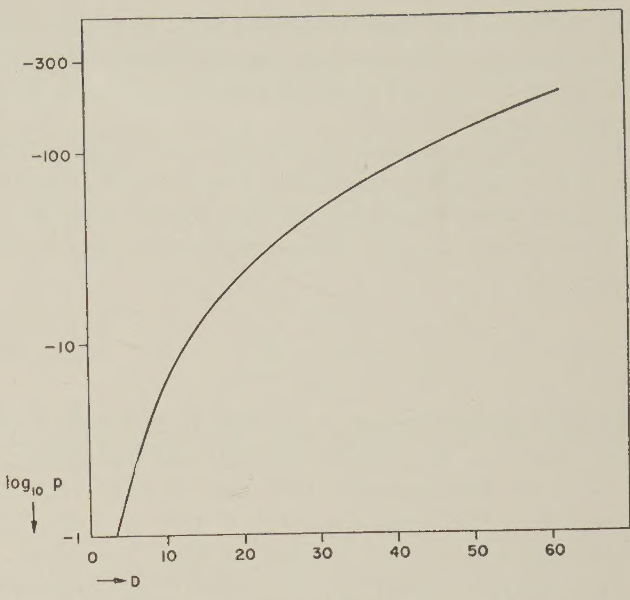


Fig. 2—The logarithm (base 10) of the binary channel error p against the signal-noise ratio D, where p and D are connected by

$$p = 2\int_{D/2}^{\infty} f(x)\, dx$$

with

$$f(x) = (1/\sqrt{2\pi}) \exp(-x^2/2).$$

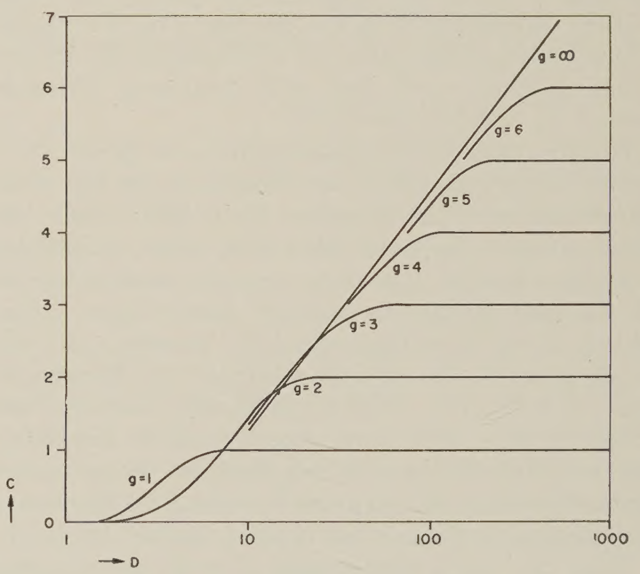


Fig. 3—Channel capacity C for various g of the "straight" multilevel channel as estimated by the "ring" model, plotted against the snr D.

The relative error

$$\epsilon = (C' - C)/C \tag{30}$$

turns out to be of the order of $-a \log a$ (if a is sufficiently small). For $D = 10$, we have $a = 3 \times 10^{-7}$ and $\epsilon = 10^{-5}$. In the case of binary channels, for small values of a, the capacity C as well as C' is very close to that of a noiseless channel, *i.e.*, unity.

For g larger than unity, in the straight channel, $P(j)$ of (7) is not necessarily unity, namely, the end-points have

larger values of $P(j)$ than the other points, since erroneous signals tend to accumulate on the endpoints. Thus, we have to use (6) instead of (8). We assume w_i in (5) to be the same for all i.

For instance, in the case of $g = 2$ we have four points, say, A, B, C, and D in their order, where A and D are the end-points. The $P(A)$ of (7) will then consist of $p_A(A) + p_B(A) + p_C(A) + p_D(A)$, where, for instance, $p_B(A)$ will be given by

$$p_B(A) = \int_{D/6}^{\infty} \tag{31}$$

since the distance between A and B is $D/3$. Actual calculation for $D = 10$ gives the straight channel capacity, $C' = 1.54$. The corresponding ring channel capacity according to (24) with the same D is $C = 1.46$. The relative error $(C' - C)/C$ is thus less than 6 per cent.

This example illustrates incidentally an important fact about the multilevel channel in general. Since the error in the binary channel is very small ($\approx 10^{-7}$), for $D = 10$, the capacity of the binary channel is practically equal to the noiseless capacity. However, dividing the interval only into $3(= 2^g - 1)$, the probability of error already becomes quite considerable; *e.g.*, $p(1) = 5 \times 10^{-2}$, resulting in a channel capacity appreciably below the noiseless capacity which is 2 in this case. This is evidently a consequence of the nature of the Gaussian distribution which decreases very fast with increasing deviation beyond the standard deviation.

We can compute the relative error $\epsilon = (C' - C)/C$ committed by the "ring" approximation for any given values of g and D by extending the method used in the above example. However, since it is rather complicated to derive a general formula, we propose to substitute for the Gaussian fluctuation another simpler type of fluctuation, for which we can easily derive an exact expression for any g and D. The capacity calculated with this simpler type of fluctuation coincides with the Gaussian case when g is very large. Furthermore, since we can obtain the exact formula for ϵ, we can learn the general nature of the dependence of ϵ on g and D, which must be more or less the same in the case of Gaussian fluctuation.

When there are a great number of levels, we can expect that the result based on the Gaussian distribution with standard deviation unity can be approximated by the use of a square distribution whose half width is of the order of unity. Thus, we propose to try a probability distribution of the following type:

$$\begin{aligned} f(x) &= 1/4 \quad \text{for} \quad -2 \le x \le 2; \\ f(x) &= 0 \qquad \text{for} \quad |x| > 2. \end{aligned} \tag{32}$$

Since $f(x)$ of (32) covers a domain of length 4, $(4/d) = (4/G)2^g$ will be the number of subintervals over which the error will extend. Therefore, if we define a parameter b by

$$b + (1/2) = 2/d = 2^{g+1}/G, \tag{33}$$

we can roughly interpret b as the number of neighboring levels on one side to which the error will extend.[5] We can now consider any pair out of three variables, g, G, and b as independent variables. For a given value of g, if we increase G, there will be a certain value of G beyond which b defined by (33) becomes negative. Therefore, we have to supplement the definition (33) by

$$b = 0 \quad \text{for} \quad G \ge 2^{g+2}. \tag{34}$$

The transition probability will be

$$p = 1/(2b + 1) \tag{35}$$

for the $(2b + 1)$ neighboring levels and will vanish beyond this domain.

The channel capacity of the ring model according to this square distribution is easily calculated with the help of (8), (33), (34), and (35):[6]

$$\begin{aligned} C &= \log G - 2 \quad \text{for} \quad G < 2^{g+2}, \\ C &= g \qquad\qquad\; \text{for} \quad G \ge 2^{g+2}. \end{aligned} \tag{36}$$

For a given value of G, if we take a large enough g such that $g > \log (G/4)$, we get $C = \log G - 2$ which agrees very nicely with the C for the Gaussian noise: $C = \log G - (1/2) \log (2 \pi e) = \log G - 2.047$ which is supposed to be correct for $g \ni \log G$. [See (20)]. At $G = 4$, C becomes zero, since the errors then cover the entire circle. Eq. (36) corresponds precisely to what was called "zeroth approximation" at the end of the third section. We can rewrite all these formulas in terms of D instead of G by the use of (23).

Now we cut the ring at one point, and assume that the "overflowing" errors in the ring model are now registered by the receiver as the signal at the end points. The points in the central part, whose number is $n - 2b$, are not affected by this modification, since the errors which reach these points in the ring model do not originate from points beyond the end-points.

The resulting channel capacity C', by the use of (6) is given by

$$\begin{aligned} C' - C = {} & \frac{2b}{n} \log (2b + 1) \\ & - \frac{(b + 2)(b + 1)}{n(2b + 1)} \log \frac{(b + 2)(b + 1)}{2} \\ & + \frac{2}{n(2b + 1)} \left(\sum_{r=1}^{b+1} - \sum_{r=b+2}^{2b} \right) r \log r. \end{aligned} \tag{37}$$

where C is the capacity of the ring model given in (36), and $n = 2^g$.

For small values of b, we can easily calculate the relative error $\epsilon = (C' - C)/C$ by this formula. Thus, for $b = 1$,

[5] To make it possible for both b and g to be integers, one would have to define b by $b \le (2/d) < b + 1$. But, this will not change the general features.

[6] Assuming b to be an integer.

we get

$$\epsilon = \frac{1.778}{D + 1.333} \Big/ \log\left(\frac{D}{4} + 0.333\right),$$

$$g = \log\left(\frac{3}{4} D + 1\right). \qquad (38)$$

and for $b = 2$

$$\epsilon = \frac{2.074}{D + 0.8} \Big/ \log\left(\frac{D}{4} + 0.2\right),$$

$$g = \log\left(\frac{5}{4} D + 1\right). \qquad (39)$$

For $b \gg 2$, it can be shown that

$$\left(\sum_{r=1}^{b+1} - \sum_{r=b+2}^{2b}\right) r \log r \approx -(b^2/2) \log (16b^2/e) \qquad (40)$$

in (37), and we obtain

$$\epsilon = \frac{2.443}{D} \Big/ \log\frac{D}{4}, \qquad g \exists \log\left(\frac{5}{4} D + 1\right). \qquad (41)$$

Fig. 4 represents ϵ as given by (41) as a function of D.

The main purpose of the use of a square distribution is to determine the behavior of ϵ as a function of g and D. We can notice from (38), (39), and (41) that ϵ does not depend appreciably on g, as has been foreseen in the preceding section. In particular, when g is larger than log $(\frac{5}{4} D + 1)$, ϵ is practically constant for a given D. We can recognize also in all three formulas (38), (39), and (41), that ϵ decreases as $1/(D \log D)$. For large values of D, all three expressions give $\epsilon \approx 2/[D \log (D/4)]$.

The numerical value of the relative error in this approximation is rather high compared with the Gaussian case. For instance, for $D = 10$, (38) (which implies $g = 3.1$ here) gives $\epsilon = 11$ per cent. For the same D, (39) ($g = 3.8$) gives $\epsilon = 13$ per cent, and (41) ($g \exists 3.8$) gives $\epsilon = 18$ per cent. In the Gaussian case for $D = 10$, $g = 2$, we had $\epsilon = 6$ per cent. However, the numerical values are not of importance here. The essential fact is that the order of magnitude of ϵ remains unchanged for a wide range of variation of g when D is once determined.

We may infer justifiably that these main features remain the same in the Gaussian case, namely, the relative error must be very insensitive to the values of g for a given value of D, and the dependence of ϵ on D must be qualitatively given by $1/(D \log D)$. Thus, we can expect that for values of D larger than 10, the relative error will rather quickly decrease with increasing D, starting from a value of the order of 6 per cent for $D = 10$, irrespective of the value of g. We can therefore conclude that the ring model is a good approximation for the problem of multilevel channels. The order of magnitude of the error committed by this model can be estimated by the method explained previously.

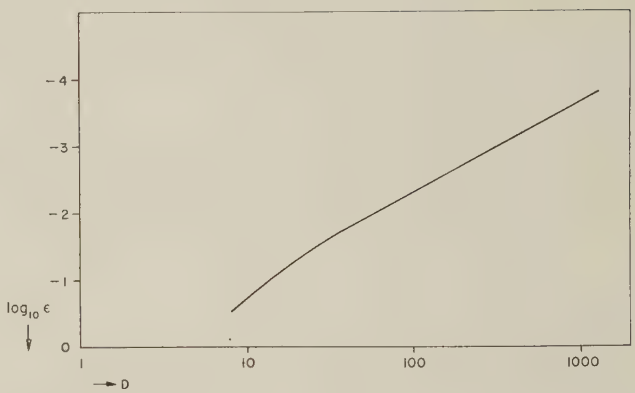


Fig. 4—The order of magnitude ($\log_{10} \epsilon$) of the fractional error ϵ committed by using the "ring" model for the "straight" multilevel channel, plotted against the snr D.

To cope with the assymmetry introduced by cutting the ring to obtain the straight multilevel channel, we can think of two methods which may lead to a certain improvement of the capacity. We may vary the weight w_i in (5) which has been assumed to be uniform, and/or vary the subintervals which have been assumed to be all equidistant. However, since the effect of the asymmetry has been shown to be very small, the change in capacity brought about by these methods will also be very small.

Conclusion

As one of the practical conclusions that can be drawn from this paper, we may repeat what has been stated at the end of the third section. Namely, when the snr is given, *i.e.*, when the probability of error of the original binary channel is given, there is a certain number of levels beyond which the channel capacity does not appreciably increase any longer by increasing the number of levels. A quantitative evaluation of such a number of levels has been given in the text.

From a theoretical point of view it is interesting to note that (16), (26), and (36) substantiate the expectation, mentioned in connection with (10), that the capacity of a multilevel channel with a very large level number will approach that of a continuous channel.

Acknowledgment

The author wishes to express his thanks to Nathaniel Rochester for drawing his attention to the problem of multilevel information channels. Appreciation is also expressed to Mr. Rochester and other members of the Information Research Division of the IBM Research Center for their stimulating discussions.

A Generalization of a Method for the Solution of the Integral Equation Arising in Optimization of Time-Varying Linear Systems with Nonstationary Inputs*

MARVIN SHINBROT†

Summary—A new method is presented for the solution of the integral equation which arises in the optimization of a system in the presence of noise when the inputs are not stationary. The method depends on the correlation functions satisfying a certain condition which, fortunately, is frequently satisfied in practical situations. A simple example is presented to illustrate the method.

Introduction

A MAJOR limitation of Wiener's [1] important theory of system optimization is the restriction to stationary inputs. Since in many practical problems—those of optimization of missile guidance systems, for example—the inputs are decidely not stationary, it would be beneficial if a general theory were developed which did not make this assumption.

The first step in this direction was taken by Booton [2] who derived an integral equation which a system with nonstationary inputs must satisfy if it is to be optimum.

Naturally, this integral equation is not of much use unless it can be solved. The extreme generality of form displayed by the equation Booton derived, however, precludes any real hope of solving it as is. Thus, the real problem which remains is one of delimiting conditions which, while mild enough to be satisfied in many practical cases, remain sufficiently restrictive for the integral equation to be solved.

Two such conditions were presented earlier [3, 4]. However, one of these two conditions had no clear physical interpretation. Consequently, it seemed desirable to see if it could be eliminated. This elimination was reported earlier [5] under the hypothesis that the noise is white. In the present paper, it will be shown how the equation can still be solved without making the offending assumption and without hypothesizing white noise.

The Integral Equation for the Optimum

In [4] the integral equation which a time-varying impulse response of a system with nonstationary inputs must satisfy, if it is to be optimum, was shown to be of the form

$$\varphi_{\mu i}(t, \tau) = \int_0^t g(t, \sigma)\varphi_{ii}(\tau, \sigma)\, d\sigma \quad \text{for} \quad 0 \leq \tau \leq t, \tag{1}$$

where $\varphi_{\mu i}$ and φ_{ii} represent correlation functions, the former of the desired output μ with the input i to the system, and the latter of the input with itself. It is the purpose of the following to solve (1) with as few conditions on these correlation functions as possible.

Assumptions

As mentioned in the Introduction, it is necessary to make some assumptions in order to be able to solve (1). The following was assumed in [3, 4]:

i) The correlation functions φ_{ii} and $\varphi_{\mu i}$ have the form

$$\varphi_{ii}(t, \tau) = \sum_{p=1}^{P} a_p(t)b_p(\tau) \tag{2}$$

$$\varphi_{\mu i}(t, \tau) = \sum_{p=1}^{P} c_p(t)b_p(\tau) \tag{3}$$

for $t \geq \tau$.

ii) If a_p and b_p have the same meaning as in assumption *i*), then the function

$$w(t, \tau) = \sum_{p} [a_p(t)b_p(\tau) - a_p(\tau)b_p(t)]$$

is a function of $t - \tau$ alone:

$$w(t, \tau) = w(t - \tau).$$

Now *i*) is very natural and will be true in many practical problems. Even if it were not, it would generally be possible to construct an approximation for which it is true. This is discussed by Shinbrot [4, 5]. On the other hand, this is not true of assumption *ii*). It is not at all clear exactly what it means physically to say that assumption *ii*) holds in any given problem. This report will present a method for solving (1) without using assumption *ii*).

Incidentally, it might be mentioned that there is no loss in generality in having the same functions $b_p(\tau)$ occurring in both (2) and (3). Suppose we had a problem with

$$\varphi_{ii} = t\tau$$

$$\varphi_{\mu i} = t^2\tau^3$$

for $t \geq \tau$. Now, φ_{ii} and $\varphi_{\mu i}$ considered as functions of τ are certainly very different. However, we can write

$$a_1(t) = t, \qquad a_2(t) = 0$$

$$b_1(\tau) = \tau, \qquad b_2(\tau) = \tau^2$$

$$c_1(t) = 0, \qquad c_2(t) = t^2$$

and with these functions a, b, c, (2) and (3) hold with $P = 2$.

One further comment is necessary before we turn to the

* Manuscript received by the PGIT, March 4, 1957.

†National Advisory Committee for Aeronautics, Ames Aeronautical Lab., Moffett Field, Calif.

solution of (1). The definition of φ_{ii} can be used to show that this function is always symmetric:

$$\varphi_{ii}(t, \tau) = \varphi_{ii}(\tau, t).$$

Hence, (2), which is true for $t \geq \tau$, implies that

$$\varphi_{ii}(t, \tau) = \sum_{p=1}^{P} a_p(\tau) b_p(t) \quad \text{for} \quad t < \tau. \tag{4}$$

In what follows, vector notation will be used. Set

$$A(t) = (a_1(t), \cdots, a_P(t))$$
$$B(t) = (b_1(t), \cdots, b_P(t))$$
$$C(t) = (c_1(t), \cdots, c_P(t)).$$

Then, assumption *i*) can be written in the alternative form (*i* bis). The correlation functions $\varphi_{ii}(t, \tau)$ have the form

$$\left.\begin{aligned} \varphi_{ii}(t, \tau) &= A(t) \cdot B(\tau) \\ \varphi_{\mu i}(t, \tau) &= C(t) \cdot B(\tau) \end{aligned}\right\} \quad \text{for} \quad t \geq \tau \tag{5}$$

where the dot denotes the ordinary scalar product of the vectors. Eq. (4) shows that (5) implies

$$\varphi_{ii}(t, \tau) = A(\tau) \cdot B(t) \quad \text{for} \quad t < \tau. \tag{6}$$

We state now that assumption (*i* bis) will be made throughout this report.

Solution of the Integral Equation

In the notation of assumption (*i* bis), (1) becomes

$$C(t) \cdot B(\tau) = A(\tau) \cdot \int_0^\tau g(t, \sigma) B(\sigma)\, d\sigma + B(\tau) \cdot \int_\tau^t g(t, \sigma) A(\sigma)\, d\sigma \quad \text{for} \quad 0 \leq \tau \leq t. \tag{7}$$

Write the second integral of (7) as $\int_0^t - \int_0^\tau$; then, after rearrangement, this equation becomes

$$\left[C(t) - \int_0^t g(t, \sigma) A(\sigma)\, d\sigma \right] \cdot B(\tau) = \int_0^\tau g(t, \sigma) [A(\tau) \cdot B(\sigma) - A(\sigma) \cdot B(\tau)]\, d\sigma \quad \text{for} \quad 0 \leq \tau \leq t. \tag{8}$$

We now attempt to write $g(t, \tau)$ in the form

$$g(t, \tau) = k(t, \tau) u(t - \tau) + h(t) \delta(t - \tau) \tag{9}$$

where $u(t)$ is the unit step function: $u(t) = 0$, $t < 0$, $u(t) = 1$, $t > 0$, and where δ denotes the Dirac δ function [7]. Note that the function $u(t - \tau)$ enters because we wish $g(t, \tau)$ to be physically realizable.

The fundamental property of the δ function that

$$\int_a^b f(\sigma) \delta(t - \sigma)\, d\sigma = \begin{cases} f(t), & \text{if} \quad a \leq t \leq b \\ 0, & \text{otherwise} \end{cases}$$

can then be used to show that (8) becomes

$$\left[D(t) - \int_0^t k(t, \sigma) A(\sigma)\, d\sigma \right] \cdot B(\tau) = \int_0^\tau k(t, \sigma) [A(\tau) \cdot B(\sigma) - A(\sigma) \cdot B(\tau)]\, d\sigma, \quad \text{for} \quad 0 \leq \tau < t \tag{10}$$

where

$$D(t) = C(t) - h(t) A(t). \tag{11}$$

Now, set

$$k(t, \tau) = G(t) \cdot \Gamma(\tau) \tag{12}$$

where G and Γ are vectors. Using (12) in (10), we see that

$$\left[D(t) - \int_0^t k(t, \sigma) A(\sigma)\, d\sigma \right] \cdot B(\tau) = G(t) \cdot \int_0^\tau \Gamma(\sigma) [A(\tau) \cdot B(\sigma) - A(\sigma) \cdot B(\tau)]\, d\sigma \quad \text{for} \quad 0 \leq \tau < t.$$

This equation is certainly true if

$$D(t) - \int_0^t k(t, \sigma) A(\sigma)\, d\sigma = G(t) \quad \text{for} \quad t > \tau \tag{13a}$$

$$B(\tau) = \int_0^\tau \Gamma(\sigma) [A(\tau) \cdot B(\sigma) - A(\sigma) \cdot B(\tau)]\, d\sigma \quad \text{for} \quad 0 \leq \tau < t. \tag{13b}$$

The solution of (1) has now been made to depend on the solution of the system (13). The process of solving (13) can be broken into three parts: the determination of Γ, of G, and, finally, of the function h of (11).

Determination of Γ

To find Γ, define the vectors

$$\left.\begin{aligned} E(\tau) &= (a_1(\tau), \cdots, a_P(\tau), b_1(\tau), \cdots, b_P(\tau)) \\ F(\tau) &= (b_1(\tau), \cdots, b_P(\tau), -a_1(\tau), \cdots, -a_P(\tau)) \end{aligned}\right\}. \tag{14}$$

In terms of these vectors, we have

$$A(\tau) \cdot B(\sigma) - A(\sigma) \cdot B(\tau) = E(\tau) \cdot F(\sigma).$$

Hence, if Γ has the components $\gamma_1, \cdots, \gamma_p$, (13b) can be written

$$b_p(\tau) = E(\tau) \cdot \int_0^\tau F(\sigma) \gamma_p(\sigma)\, d\sigma \quad \text{for} \quad 0 \leq \tau < t, p = 1, \cdots, P,$$

or

$$b_p(\tau) = \sum_{q=1}^{2P} e_q(\tau) \int_0^\tau f_q(\sigma) \gamma_p(\sigma)\, d\sigma \quad \text{for} \quad 0 \leq \tau < t, p = 1, \cdots, P$$

where

$$e_q(\tau) = a_q(\tau)$$
$$f_q(\tau) = b_q(\tau)$$

for $q = 1, \cdots, P$, and

$$e_q(\tau) = b_{q-P}(\tau)$$
$$f_q(\tau) = -a_{q-P}(\tau)$$

for $q = P + 1, \cdots, 2P$.

Now, it is entirely possible that the functions $e_q(\tau)$ are not all linearly independent. In this case, certain of the terms in the sum on the right side of (21) can be collected together until an equation of the form

$$b_p(\tau) = \sum_{q=1}^{Q} \epsilon_q(\tau) \int_0^\tau \varphi_q(\sigma)\gamma_p(\sigma)\, d\sigma \quad \text{for} \quad 0 \le \tau < t, p = 1, \cdots, P \tag{16}$$

is obtained where the functions ϵ_q *are* linearly independent. This process will be illustrated in the last section. Eq. (16) can be reduced immediately to a system of differential equations. In fact, differentiating (16) r times gives

$$\sum_{q=1}^{Q} \epsilon_q^{(r)}(\tau) \int_0^\tau \varphi_q(\sigma)\gamma_p(\sigma)\, d\sigma = b_p^{(r)}(\tau) - \sum_{s=1}^{r} \binom{r}{s} \sum_{q=1}^{Q} \epsilon_q^{(r-s)}(\tau)[\varphi_q(\tau)\gamma_p(\tau)]^{(s-1)} \quad \text{for} \quad 0 \le \tau < t, p = 1, \cdots, P \tag{17}$$

where $\binom{r}{s}$ denotes the binomial coefficient:

$$\binom{r}{s} = \frac{r!}{s!(r-s)!}.$$

Eq. (17) with $r = 0, 1, \cdots, Q - 1$ represents Q simultaneous equations in the Q unknowns $\int_0^\tau \varphi_q(\sigma)\gamma_p(\sigma)d\sigma$, $q = 1, \cdots, Q$ (p fixed). Furthermore, the coefficient determinant

$$\det (\epsilon_q^{(r)}) \tag{18}$$

is never zero, since we know the functions ϵ_q to be independent, from which it follows that their Wronskian (18) is different from zero. Consequently, (17) can be solved for the integrals

$$\int_0^\tau \varphi_q(\sigma)\gamma_p(\sigma)\, d\sigma$$

in terms of the functions $\gamma_p(\tau)$ and their derivatives. Differentiating these solution equations once more results in a linear differential equation for $\gamma_p(\tau)$.

This of course still does not determine the γ_p explicitly, for it can easily happen that the differential equations may not be explicitly soluble. If they are (and this frequently happens in practice), we can now go on with the solution, but even if they are not, much is known about finding approximate solutions of differential equations—far more than is known about approximations to solutions of integral equations.

The functions γ_p have now been determined with the exception of certain constants which occur when the differential equation for γ_p is solved. These constants can be determined by substitution of the expression γ_p into the original integral equation (13b).

Determination of G

Thus, γ_p can be found completely. To find the components $g_p(t)$ of the vector $G(t)$ occurring in (12), consider (13a). Use of (12) in (13a) gives

$$d_p(t) = g_p(t) + \sum_{q=1}^{Q} g_q(t) \int_0^t \gamma_q(\sigma)a_p(\sigma)\, d\sigma \quad \text{for} \quad t > \tau, p = 1, \cdots, Q. \tag{19}$$

Since the functions γ_q are now known, (19) can be considered as Q simultaneous equations to be solved for the Q unknowns $g_p(t)$ as functions of $h(t)$ since, according to (11), $d_p(t)$ depends on $h(t)$.

Determination of h

It remains to show how h may be found. The functions $g_p(t)$ have been determined in terms of the functions $d_p(t)$; however, these functions are not completely known since the function $h(t)$ occurring in (11) has not as yet been specified. On the other hand, it can be seen from what has gone before that whatever this function h may be, the function (9) with $k(t, \tau)$ being given by (12) and the vectors G and Γ being determined by the method just described satisfies the fundamental integral equation (1). Consequently, we may say that we have found an infinite number of solutions of (1). How, then, do we specify the function h?

To answer this, consider the functions $\gamma_p(\tau)$. It can be seen that they have been determined independently of h. Also, since the functions $b_p(\tau)$ of (16) are zero for $t < 0$, it can be seen that the differentiated equation (17) will generally include some derivatives of δ functions on the right-hand side. In fact, assuming that the functions b_p themselves do not involve δ functions, it can be seen that the differential equation for γ_p will in general be of order $Q - 2$. On the other hand, the Qth derivative of b_p will occur in it, which means, since $b_p(\tau)$ may be discontinuous at $\tau = 0$, that $\delta^{(Q-1)}(\tau)$ will probably occur in the differential equation. Hence, the solution for $\gamma_p(\tau)$ generally will depend linearly on the first derivative of $\delta(\tau)$. This means that the optimum system may involve a differentiator. Since it is difficult, if not impossible, to differentiate a noisy input, it is highly desirable that this term $\dot{\delta}(\tau)$ be eliminated. The function $h(t)$ can be used for this purpose.

To see this, consider (11) and (19). From these equations, it follows that the $g_p(t)$ depend linearly on $h(t)$. Hence, from (12), the $\dot{\delta}(\tau)$ terms which occur in $\Gamma(\tau)$ will be multiplied by a linear function of $h(t)$; *i.e.*, $k(t, \tau)$ will contain a term of the form $[\alpha(t) + \beta(t)h(t)]\dot{\delta}(\tau)$ and, furthermore, $\dot{\delta}(\tau)$ will appear nowhere else in the expression for k. Thus if $h(t)$ is set equal to $-\alpha(t)/\beta(t)$, the optimum will not involve any differentiators. This process will be illustrated in the next section.

Example

To illustrate the method of the preceding section, we now present a detailed example.

Consider the problem of determining the position of a moving particle when the measurements are corrupted by noise. In order to simplify the problem to the point where the method will not be obscured by the details of

he calculation, we make the following assumptions:[1]

a) The particle moves along a given straight line with constant (but unknown) speed.
b) The particle can come from only one direction along this line.
c) The time when the particle first appears—at the limit of our measuring devices, say—is known.
d) The noise and the particle position (message) are independent.
e) The noise is stationary and is described by an auto-correlation function of the form

$$\varphi_{nn}(t, \tau) = \lambda e^{-\beta|t-\tau|}. \tag{20}$$

Define a coordinate system with one axis along the line in which the particle moves and with origin at the point of its appearance. Then, under the assumptions made, a typical message will be

$$m(t; \alpha) = \alpha t$$

where α is the speed of the particle. Hence, we can compute

$$\begin{aligned}\varphi_{mm}(t, \tau) &= Av\langle \alpha t \cdot \alpha \tau \rangle \\ &= \overline{\alpha^2} t\tau\end{aligned}$$

where $\overline{\alpha^2}$ is the expected mean-square velocity of the particle. We assume an approximation to $\overline{\alpha^2}$ is known.

By d) and e), we have

$$\begin{aligned}\varphi_{mi}(t, \tau) &= \varphi_{mm}(t, \tau) \\ &= \overline{\alpha^2} t\tau\end{aligned}$$

while from e),

$$\begin{aligned}\varphi_{ii}(t, \tau) &= \varphi_{mm}(t, \tau) + \varphi_{nn}(t, \tau) \\ &= \overline{\alpha^2} t\tau + \lambda e^{-\beta|t-\tau|}.\end{aligned}$$

Since the problem is one of filtering, the desired output μ is the message m. Thus, in notation of (i bis), we can set

$$\left.\begin{aligned} A(t) &= (\overline{\alpha^2} t, \lambda e^{-\beta t}) && (21a) \\ B(t) &= (t, e^{\beta t}) && (21b) \\ C(t) &= (\overline{\alpha^2} t, 0) && (21c) \end{aligned}\right\}$$

for $t \geq 0$.[2] Thus, from (14),

$$E(\tau) = (\overline{\alpha^2}\tau, \lambda e^{-\beta\tau}, \tau, e^{\beta\tau})$$

$$F(\tau) = (\tau, e^{\beta\tau}, -\overline{\alpha^2}\tau, -\lambda e^{-\beta\tau}).$$

Therefore, (15) becomes

$$\begin{aligned} b_p(\tau) = \overline{\alpha^2}\tau \int_0^\tau \sigma\gamma_p(\sigma)\, d\sigma + \lambda e^{-\beta\tau} \int_0^\tau e^{\beta\sigma}\gamma_p(\sigma)\, d\sigma \\ - \overline{\alpha^2}\tau \int_0^\tau \sigma\gamma_p(\sigma)\, d\sigma - \lambda e^{\beta\tau} \int_0^\tau e^{-\beta\sigma}\gamma_p(\sigma)\, d\sigma \\ \text{for} \quad 0 \leq \tau < t. \end{aligned} \tag{22}$$

[1] All of these assumptions can be eliminated.

[2] Note, incidentally, that for this simple example, assumption (2) is fulfilled and so the simpler method of Shinbrot [3, 4] can also be used to solve it.

Now, clearly, the components of $E(\tau)$ are not linearly independent. This fact manifests itself in (22) by the cancellation of the first and third integrals on the right-hand side. Thus, (16) is

$$b_p(\tau) = \lambda e^{-\beta\tau} \int_0^\tau e^{\beta\sigma}\gamma_p(\sigma)\, d\sigma - \lambda e^{\beta\tau} \int_0^\tau e^{-\beta\sigma}\gamma_p(\sigma)\, d\sigma \quad \text{for} \quad 0 \leq \tau < t. \tag{23}$$

Differentiate this equation. We obtain

$$-\frac{1}{\beta}\dot{b}_p(\tau) = \lambda e^{-\beta\tau} \int_0^\tau e^{\beta\sigma}\gamma_p(\sigma)\, d\sigma + \lambda e^{\beta\tau} \int_0^\tau e^{-\beta\sigma}\gamma_p(\sigma)\, d\sigma \quad \text{for} \quad 0 \leq \tau < t. \tag{24}$$

Adding the two preceding equations results in

$$\int_0^\tau e^{\beta\sigma}\gamma_p(\sigma)\, d\sigma = \frac{e^{\beta\tau}}{2\lambda}\left[b_p(\tau) - \frac{1}{\beta}\dot{b}_p(\tau)\right],$$

so that

$$\gamma_p(\tau) = \frac{\beta^2 b_p(\tau) - \ddot{b}_p(\tau)}{2\beta\lambda}. \tag{25}$$

It is important to note that the same result would have been obtained had we, instead of adding (23) and (24), subtracted them to get an expression for $\int_0^\tau e^{-\beta\sigma}\gamma_p(\sigma)d\sigma$.

We now write, from (21) and the fact that $B(\tau) = 0$ for $\tau < 0$,

$$b_1(\tau) = \tau u(\tau)$$

$$b_2(\tau) = e^{\beta\tau}u(\tau)$$

where $u(\tau)$ is the unit step. Hence, from (25),

$$\left.\begin{aligned} \gamma_1(\tau) &= \frac{\beta^2\tau u(\tau) - \delta(\tau)}{2\beta\lambda} \\ \gamma_2(\tau) &= -\frac{\dot{\delta}(\tau) + 2\beta\delta(\tau)}{2\beta\lambda} e^{\beta\tau} \end{aligned}\right\}. \tag{26}$$

We now turn to the computation of the functions $g_p(t)$. From (21a) and (26), we find

$$\int_0^t a_1(\tau)\gamma_1(\tau)\, d\tau = \frac{\beta\overline{\alpha^2}}{6\lambda} t^3$$

$$\int_0^t a_2(\tau)\gamma_1(\tau)\, d\tau = -\frac{1 + \beta t}{2\beta} e^{-\beta t}$$

$$\int_0^t a_1(\tau)\gamma_2(\tau)\, d\tau = \frac{\overline{\alpha^2}}{2\beta\lambda}$$

$$\int_0^t a_2(\tau)\gamma_2(\tau)\, d\tau = -1.$$

In the third of these equations, we have used the fact that

$$\int_0^t f(\tau)\dot{\delta}(t - \tau)\, d\tau = \dot{f}(t)$$

to prove that

$$\int_0^t f(\tau)\dot{\delta}(\tau)\,d\tau = -\dot{f}(0)).$$

Consequently, (19) becomes

$$\left(1 + \frac{\beta\overline{\alpha^2}}{6\lambda}t^3\right)g_1(t) + \frac{\overline{\alpha^2}}{2\beta\lambda}g_2(t) = d_1(t)$$

$$-\frac{1+\beta t}{2\beta}e^{-\beta t}g_1(t) = d_2(t).$$

Hence,

$$\left.\begin{aligned} g_1(t) &= -\frac{2\beta e^{\beta t}\,d_2(t)}{1+\beta t} \\ g_2(t) &= \frac{2\beta\lambda}{\overline{\alpha^2}}d_1(t) + \frac{2\beta^2}{3\overline{\alpha^2}}\cdot\frac{6\lambda + \beta\overline{\alpha^2}t^3}{1+\beta t}e^{\beta t}\,d_2(t) \end{aligned}\right\}. \tag{27}$$

Now, from (11) and (21),

$$\left.\begin{aligned} d_1(t) &= \overline{\alpha^2}t[1-h(t)] \\ d_2(t) &= -\lambda h(t)e^{-\beta t} \end{aligned}\right\}. \tag{28}$$

Notice that $\gamma_1(\tau)$ does not contain the derivative of a δ function. Thus, the optimum impulse response will not involve a differentiator if and only if $g_2(t)$— which multiplies $\gamma_2(\tau)$ in the expression for $g(t, \tau)$—is zero. Setting $g_2(t)$ from (27) equal to zero and using (28) gives

$$h(t) = \frac{3\overline{\alpha^2}t(1+\beta t)}{6\beta\lambda + 3\overline{\alpha^2}t + 3\beta\overline{\alpha^2}t^2 + \beta^2\overline{\alpha^2}t^3}. \tag{29}$$

With this function $h(t)$, substitution of (28) in (27) gives

$$g_1(t) = \frac{2\beta\lambda}{1+\beta t}h(t)$$

$$g_2(t) = 0.$$

Hence, from (9) and (12), the optimum impulse response is

$$g(t,\tau) = \frac{h(t)}{1+\beta t}[\beta^2\tau u(t-\tau) + (1+\beta t)\delta(t-\tau) - \delta(\tau)] \tag{30}$$

where $h(t)$ is given by (29).

For this example, the minimum mean-square error can be found to be

$$E^2_{\min}(t) = \overline{\alpha^2}t^2 - \overline{\alpha^2}t\int_0^t \tau g(t,\tau)\,d\tau$$

$$= \frac{6\beta\lambda\overline{\alpha^2}t^2}{6\beta\lambda + 3\overline{\alpha^2}t + 3\beta\overline{\alpha^2}t^2 + \beta^2\overline{\alpha^2}t^3}$$

$$\sim \frac{6\lambda}{\beta t} \quad \text{as} \quad t \to \infty.$$

Note that this says that by waiting long enough, the approximation to the particle's position can be made as good as desired.

The impulse response (30) can be interpreted in terms of a transfer function. This problem is considered elsewhere [4].

Bibliography

[1] Wiener, N. *Extrapolation, Interpolation, and Smoothing of Stationary Time Series*, New York: John Wiley and Sons, Inc., 1949.

[2] Booton, R. C., Jr. "An Optimization Theory for Time-Varying Linear Systems with Nonstationary Statistical Inputs," Dynamic Analysis and Control Laboratory, Massachusetts Institute of Technology, Cambridge, Mass., Meteor Report 72; July, 1951.

[3] Shinbrot, M. "On the Integral Equation Occurring in Optimization Theory With Nonstationary Inputs," *Journal of Mathematics and Physics*, vol. 36 (1957), pp. 121-129.

[4] ———. "On a Method for Optimization of Time-Varying Linear Systems with Nonstationary Inputs," National Advisory Committee for Aeronautics Technical Note 3791; 1956.

[5] ———. "Optimization of Time-Varying Linear Systems With Nonstationary Inputs," to appear in the *Transactions of the American Society of Mechanical Engineers*.

[6] Levinson, N. "A Heuristic Exposition of Wiener's Mathematical Theory of Prediction and Filtering," *Journal of Mathematics and Physics*, Vol. 25 (1946), pp. 261-278.

[7] Dirac, P. A. M. *The Principles of Quantum Mechanics*, Oxford: Clarendon Press, third ed., 1947.

On the Mean Square Noise Power of an Optimum Linear Discrete Filter Operating on Polynomial Plus White Noise Input*

MARVIN BLUM†

Summary—In a recent article[1] Johnson presents an asymptotic formula for the output noise power of an optimum filter designed to make a zero-lag estimate of either the input or its derivatives. It is assumed that the input function consists of a nonrandom polynomial plus stationary uncorrelated noise.

It is the purpose of this paper to present an exact formula for the output noise power for the same input model. The formula presented is more general in that the estimation can be for any lag α with respect to the latest data point.

Tables and graphs of the root mean square error for the zero-lag estimation of the 0th, 1st, and 2nd derivative are presented as a function of the input polynomial up to degree 5 and memory spans up to 100 sample points. A comparison is made of the relative error in root mean square using the asymptotic formula derived by Johnson.

Introduction

In a paper on the same topic by the author,[2] a generalization of the problem of determining the ordinates of the weighting sequence of the optimum linear discrete filter is considered. The special case for the polynomial plus white noise is considered and the formula for the weighting sequence is obtained in (49) to (53). An equivalence between the optimum filtering problem and minimum variance curve fitting techniques is proven. For the purposes of simplicity of derivation, it will be simpler to utilize the concepts of least squares curve fitting since the weighting sequence is invariant under a time translation and may be determined over a standard interval.

Analysis

Consider a set of equally spaced data points (u, y_u) $u = 1, 2, \cdots, M$. The problem is to fit a least squares polynomial of degree n to these points and to estimate the Kth derivative of the observed data from the curve fit at any point on the u scale.

For the purpose of the analysis it is convenient to utilize orthogonal polynomials in the curve fitting procedure. Thus let the true polynomial be given by

$$P(u) = \sum_{L=0}^{n} a_L \xi_L(u), \qquad u = 1, 2, \cdots, M \tag{1}$$

where the polynomials $\xi_L(u)$ are orthogonal,[3] *e.g.*, satisfy the following relationships

$$\sum_{u=1}^{M} \xi_h(u)\xi_L(u) = 0, \qquad h \neq L, \tag{2}$$

$$\sum_{u=1}^{M} \xi_L^2(u) = S(L, M). \tag{3}$$

It is assumed that the observations y_u are given by

$$y_u = P(u) + N(u).$$

The $N(u)$ are assumed to be random, stationary, and uncorrelated errors.

Then the least squares estimates $\hat{a}_L$ of the coefficients a_L are obtained by minimizing

$$I = \sum_{u=1}^{M} \left[\sum_{L=0}^{n} \hat{a}_L \xi_L(u) - y_u \right]^2 \tag{4}$$

with respect to each of the parameters $\hat{a}_1$.

Thus one obtains

$$\frac{\partial I}{\partial \hat{a}_L} = \sum_{u=1}^{M} 2\left[\sum_{v=0}^{n} \hat{a}_v \xi_v(u) - y_u \right] \xi_L(u) = 0, \qquad L = 0, 1, \cdots, n. \tag{5}$$

Solving (5) for $\hat{a}_L$ one obtains

$$\hat{a}_L = \sum_{u=1}^{M} \frac{y_u \xi_L(u)}{S(L, M)}, \qquad L = 0, 1, \cdots, n. \tag{6}$$

By substituting (6) one obtains the curve fit relationship

$$Y(u) = \sum_{L=0}^{n} \hat{a}_L \xi_L(u). \tag{7}$$

To evaluate the estimate of the Kth derivative at $u = M + \alpha$ one need only take the Kth derivative of both sides of (7), (considering u as a continuous variable) and obtain

$$\left.\frac{d^K Y(u)}{du^K}\right|_{u=M+\alpha} \equiv Y_{(M+\alpha)}^{(K)} = \sum_{L=K}^{n} \hat{a}_L \left.\frac{d^K}{du^K} \xi_L(u)\right|_{u=M+\alpha}. \tag{8}$$

Let

$$\left.\frac{d^K}{du^K} \xi_L(u)\right|_{u=M+\alpha} \equiv \xi_L^{(K)}(M + \alpha), \tag{9}$$

* Manuscript received by the PGIT, March 22, 1957.

† Convair, San Diego, Calif.

[1] K. R. Johnson, "Optimum, linear, discrete filterings of signals containing a nonrandom component," IRE Trans., vol. IT-2, pp. 49-55; June, 1956.

[2] M. Blum, "An extension of the minimum mean square prediction theory for sampled input signals," IRE Trans., vol. IT-2, pp. 176-184; September, 1956.

[3] R. A. Fischer and F. Yates, "Statistical Tables for Biological Agricultural and Medical Research," Oliver and Boyd, Edinburgh, Scotland; 1938.

and substituting (6) into (8) one obtains

$$Y^{(K)}(M+\alpha) = \sum_{L=K}^{n} \sum_{u=1}^{M} \frac{y_u \xi_L(u) \xi_L^{(K)}(M+\alpha)}{S(L, M)}. \quad (10)$$

Let

$$W_{M-u} = \sum_{L=K}^{n} \frac{\xi_L(u)\xi_L^{(K)}(M+\alpha)}{S(L, M)}, \quad u = 1, 2, \cdots, M \quad (11)$$

then

$$Y_{M+j}^{(K)}(M+\alpha+j) = \sum_{u=1}^{M} W_{M-u} y_{u+j},$$

$$j = 0 \pm 1 \pm 2 \cdots. \quad (12)$$

Eq. (12) is directly interpreted as the input-output relationship of a digital filter with weighting sequence $W_0, W_1, \cdots, W_{M-1}$. The input is the sequence y_{u+j} and the output is Y_{M+j} $(M + j + \alpha)$. The output is available in real time after the last data point is sampled and estimates the Kth derivative of the input at $u = M + \alpha + j$. The filter has a finite memory over the interval $(M - 1)T$.

Since the estimators $\hat{a}_L$ are unbiased, the error in estimate is given by $(j = 0)$,

$$\Delta = \left[Y_M^{(K)}(M+\alpha) - \sum_{L=K}^{n} a_L \xi_L^{(K)}(M+\alpha) \right]$$

$$\Delta = \sum_{u=1}^{M} W_{M-u} N(u). \quad (13)$$

The mean square error of estimate is given by

$$\sigma_\Delta^2 = \sigma_N^2 \sum_{u=1}^{M} [W_{M-u}]^2$$

$$(\sigma_N^2 = \text{noise mean square error}). \quad (14)$$

Substituting (11), (2), and (3) into (14) one obtains,

$$\frac{\sigma_\Delta^2}{\sigma_N^2} \equiv \delta^2(M, \alpha, K, n) = \sum_{L=K}^{n} \frac{[\xi_L^{(K)}(M+\alpha)]^2}{S(L, M)} \quad (15)$$

which is the main result of this paper.

Eq. (15) has been derived for unit time between samples. If the interval between samples is given by T then (11) is modified as follows:

$$W_{M-u,T} = \frac{1}{(T)^K} W_{M-u}, \quad (16)$$

and (15) becomes

$$\frac{\sigma_\Delta^2 (T)^{2K}}{\sigma_N^2} = \delta^2(M, \alpha, K, n). \quad (17)$$

Note that

$$\delta^2(M, \alpha, K, n) = \delta^2(M, \alpha, K, n-1) + \frac{[\xi_n^{(K)}(M+\alpha)]^2}{S(n, M)}, \quad (18)$$

so that increasing the degree of curve fit is never associated with a decrease in σ_Δ^2 since the second term of (18) is positive definite for fixed M, α, and K.

Special Case

Special formulas for δ^2 are as follows. Let $K = 0$, $\alpha = 0, -1, -2, \cdots, (M - 1)$, then

$$\delta^2(M, \alpha, 0, n) = W_{-\alpha}.$$

As an example, when $\alpha = 0$, $K = 0$, one obtains a zero-lag estimate of the input. The mean square error output is then proportional to W_0, the coefficient which multiplies the latest data point, *e.g.*,

$$\sigma_\Delta^2 = \sigma_N^2 \cdot W_0.$$

Other relationships on the δ which may be useful are as follows. Let the order of the derivative equal the order of degree of curve fit, *e.g.*, $K = n$ then

$$\delta^2(M, \alpha, n, n) = \frac{(n!)^2}{S(n, M)} \quad (18)$$

and is independent of α.

Let the order of the derivative equal one less than the degree of the curve fitting polynomial, *e.g.*, $K = n - 1$, then

$$\delta^2(M, \alpha, n-1, n) = \delta^2(M, \alpha, n-1, n-1) + \left[\frac{M + 2\alpha - 1}{2}\right]^2 \delta^2(M, \alpha, n, n) \quad (19)$$

$$= \frac{[(n-1)!]^2}{S(n-1, M)} + \left[\frac{M + 2\alpha - 1}{2}\right]^2 \frac{[n!]^2}{S(n, M)}. \quad (20)$$

When $\alpha = -[(M - 1)/2]$, *e.g.*, the midpoint of the curve fitting interval,

$$\delta\left(M, \alpha = -\left[\frac{M-1}{2}\right], n-1, n\right)$$

$$= \delta^2(M, \alpha, n-1, n-1). \quad (21)$$

This represents the minimum δ^2 obtainable with respect to α.

For $\alpha \ll M$, the asymptotic value of δ is given by

$$\delta(M, \alpha, K, n)$$

$$\cong \frac{\sqrt{2K+1}(2K)! \sum_{j=0}^{2K+1} \binom{2K+1}{j}\binom{n-K}{j}}{(M)^{(2k+1)/2}(K)!},$$

$$K = 0, 1, 2, \cdots \qquad n = K, K+1, \cdots. \quad (22)$$

Tables I and II, and Table III, p. 228, present the exact values of δ using (17) and the percentage relative error in δ using (22). Figs. 1 and 2, p. 229, and Fig. 3, p. 230, present a plot of δ using (15) for $M = 10$ to 100 for purposes of interpolation.

Eq. (17) is identical with the results one would obtain from Blum[2] as are the values of the weighting sequence.

The interpretation of the parameter α is as follows: when $\alpha = 0$, one obtains a zero-lag estimate with respect to the latest data point, when $-(M - u \rightarrow 1) < \alpha > 0$ one obtains an extrapolation, and when $(M - 1) < \alpha < 0$ one

TABLE I

Table of $R(M, \alpha, K, n)$ fixed α for evaluating the relative errors using the asymptotic root mean square formula compared with zero-lag $\alpha = 0)$ estimation of the input $(K = 0)$ as a function of the degree of the curve fitting polynomials (n) and the number of data points (M).

$$R = \left[\frac{\text{asymptotic } \delta - \text{exact } \delta}{\text{exact } \delta}\right] \times 100 = \text{per cent error in } \delta \text{ using asymptotic formula.}$$

Table of $\delta(M, \alpha, K, n)$, $\alpha = 0$, $K = 0$ for evaluating the root mean square error for zero-lag $(\alpha = 0)$ estimation of the input $(K = 0)$ s a function of the degree of the curve fitting polynomial (n) and the number of data points (M).

$M \diagdown n$		0	1	2	3	4	5
2	δ	0.70711					
	R	0					
3	δ	0.57735	0.91287				
	R	0	26.5				
4	δ	0.50000	0.83666	0.97468			
	R	0	19.5	53.9			
5	δ	0.44721	0.77460	0.94112	0.99283		
	R	0	15.5	42.6	80.2		
6	δ	0.40825	0.72375	0.90633	0.97996	0.99801	
	R	0	12.8	35.1	66.6	104.5	
7	δ	0.37796	0.68139	0.87287	0.96362	0.99348	0.99946
	R	0	10.9	29.9	56.9	90.2	126.9
8	δ	0.35355	0.64550	0.84162	0.94548	0.98665	0.99796
	R	0	9.54	26.0	49.6	79.2	112.6
9	δ	0.33333	0.61464	0.81278	0.92660	0.97800	0.99533
	R	0	8.46	23.0	43.9	70.4	100.9
10	δ	0.31623	0.58775	0.78625	0.90762	0.96802	0.99157
	R	0	7.61	20.7	39.4	63.3	91.3
20	δ	0.22361	0.43095	0.60892	0.74985	0.85231	0.92022
	R	0	3.77	10.1	19.3	31.1	45.8
50	δ	0.14142	0.27865	0.40784	0.52578	0.63011	0.71946
	R	0	1.50	4.02	7.59	12.2	17.9
95	δ	0.10260	0.20359	0.30142	0.39471	0.48223	0.56299
	R	0	0.790	2.11	3.97	6.38	9.34
1001	δ	0.031607	0.063167	0.094632	0.12596	0.15709	0.18800
	R	0	0.0749	0.1996	0.374	0.5998	0.87517

TABLE II

Table of $R(M, \alpha, K, n)$ fixed α for evaluating the relative errors using the asymptotic root mean square formula compared with zero-lag $(\alpha = 0)$ estimation of the first derivative of the input $(K = 1)$ as a function of the degree of the curve fitting polynomials (n) and the number of data points (M).

$$R = \left[\frac{\text{asymptotic } \delta - \text{exact } \delta}{\text{exact } \delta}\right] \times 100 = \text{per cent error in } \delta \text{ using asymptotic formula.}$$

Table of $\delta(M, \alpha, K, n)$, $\alpha = 0$, $K = 1$ for evaluating the root mean square error for zero-lag $(\alpha = 0)$ estimation of the first derivative of the input $(K = 1)$ as a function of the degree of the curve fitting polynomial (n) and the number of data points (M).

$M \diagdown n \rightarrow$		1	2	3	4	5
2	δ	1.0				
	R	22.5				
3	δ	0.70711	2.5495			
	R	−5.72	4.59			
4	δ	0.44721	1.5652	3.83695		
	R	−3.17	10.7	12.8		
5	δ	0.31623	1.1148	2.52566	5.5839	
	R	−2.02	11.2	22.7	10.9	
6	δ	0.23905	0.85252	1.90348	3.6802	8.2418
	R	−1.40	10.6	23.8	28.1	0.0944
7	δ	0.18898	0.68138	1.52189	2.8226	5.2190
	R	−1.03	9.80	22.9	32.5	25.4
8	δ	0.15430	0.56167	1.26041	2.3028	3.9614
	R	−0.784	9.03	21.5	33.0	35.3
9	δ	0.12910	0.47377	1.06943	1.9445	3.2351
	R	−0.619	8.32	20.0	31.9	38.8
10	δ	0.11010	0.40685	0.92389	1.6794	2.7476
	R	−0.501	7.70	18.6	30.5	39.5
20	δ	0.038778	0.14855	0.35079	0.65608	1.0684
	R	−0.125	4.29	10.4	18.1	26.9
50	δ	0.0097999	0.038494	0.093871	0.18201	0.30714
	R	−0.020	1.81	4.37	7.66	11.7
95	δ	0.0037414	0.014821	0.036558	0.071883	0.12326
	R	−0.00553	0.970	2.34	4.09	6.23
1001	δ	0.000109380	0.00043711	0.0010913	0.0021790	0.0038054
	R	0.0....0	0.0937	0.225	0.394	0.602

TABLE III

Table of $R(M, \alpha, K, n)$ fixed α for evaluating the relative errors using the asymptotic root mean square formula compared with zero-lag ($\alpha = 0$) estimation of the second derivative of the input ($K = 2$) as a function of the degree of the curve fitting polynomials (n) and the number of data points (M).

$$R = \left[\frac{\text{asymptotic } \delta - \text{exact } \delta}{\text{exact } \delta}\right] \times 100 = \text{per cent error in } \delta \text{ using asymptotic formula.}$$

Table of $\delta(M, \alpha, K, n)$ $\alpha = 0$, $K = 2$ for evaluating the root mean square error for zero-lag ($\alpha = 0$) estimation of the second derivative of the input ($K = 2$) as a function of the degree of the curve fitting polynomial (n) and the number of data points (M).

$M \diagdown n \rightarrow$		2	3	4	5
3	δ	2.4495			
	R	−29.7			
4	δ	1.0	6.7823		
	R	−16.1	−25.8		
5	δ	0.53452	3.2071	14.017	
	R	−10.2	−10.2	−28.1	
6	δ	0.32733	1.8919	7.0312	26.312
	R	−7.04	−3.50	−9.12	−35.2
7	δ	0.21822	1.2440	4.3582	13.334
	R	−5.15	−0.175	−0.268	−13.1
8	δ	0.15430	0.87535	2.9852	8.4113
	R	−3.93	+1.60	0.428	−1.31
9	δ	0.11396	0.64578	2.1729	5.8714
	R	−3.10	2.60	0.672	5.32
10	δ	0.087039	0.49355	1.6494	4.3527
	R	−2.51	3.15	8.03	9.17
20	δ	0.015094	0.087123	0.29179	0.74414
	R	−0.626	3.30	7.95	12.9
50	δ	0.0015194	0.0089549	0.030657	0.079685
	R	−0.100	1.70	3.97	6.67
95	δ	0.00030512	0.0018129	0.0062672	0.016468
	R	−0.0277	0.957	2.21	3.73
1001	δ	0.84641×10^{-6}	0.50735×10^{-5}	0.17741×10^{-4}	0.47264×10^{-4}
	R	0.0....0	0.0972	0.191	0.285

obtains interpolation of the input polynomial. A more detailed discussion of σ_Δ^2 as a function of α is available.[2]

Appendix I contains a summary of a few useful properties of the orthogonal polynomials, while Appendix II contains a discussion of (22).

Conclusion

An exact equation for the mean square error of the output of an optimum digital filter has been presented. The formula was derived using curve fitting concepts to demonstrate the relationship between the concepts of parameter estimation in curve fitting and weighting function optimization in linear filtering.

Eq. (11) represents a convenient formula for computing the weighting sequence of the digital filter.

From Tables I-III one may determine both the relative error associated with using the asymptotic formula for δ and the exact value of δ for small M.

In Figs. 1–3 the values of δ can be determined for those values of M not tabulated. For $M > 100$ one can either extrapolate linearly on log-log paper or use the asymptotic formula.

Appendix I

A listing of the orthogonal polynomials in consistent notation is as follows:[4]

$$
\begin{aligned}
\xi_0(u) &= 1 \\
\xi_1(u) &= (u - \bar{u}), \ \bar{u} = \frac{M+1}{2} \\
\xi_2(u) &= (u - \bar{u})^2 - \frac{M^2 - 1}{12} \\
\xi_3(u) &= (u - \bar{u})^3 - (u - \bar{u})\left[\frac{3M^2 - 7}{20}\right] \\
\xi_4(u) &= (u - \bar{u})^4 - (u - \bar{u})^2\left[\frac{3M^2 - 13}{14}\right] + \frac{3(M^2 - 1)(M^2 - 9)}{560} \\
\xi_5(u) &= (u - \bar{u})^5 - (u - \bar{u})^3\left[\frac{5(M^2 - 7)}{18}\right] + (u - \bar{u})\left[\frac{15M^4 - 230M^2 + 407}{1008}\right].
\end{aligned} \tag{23}
$$

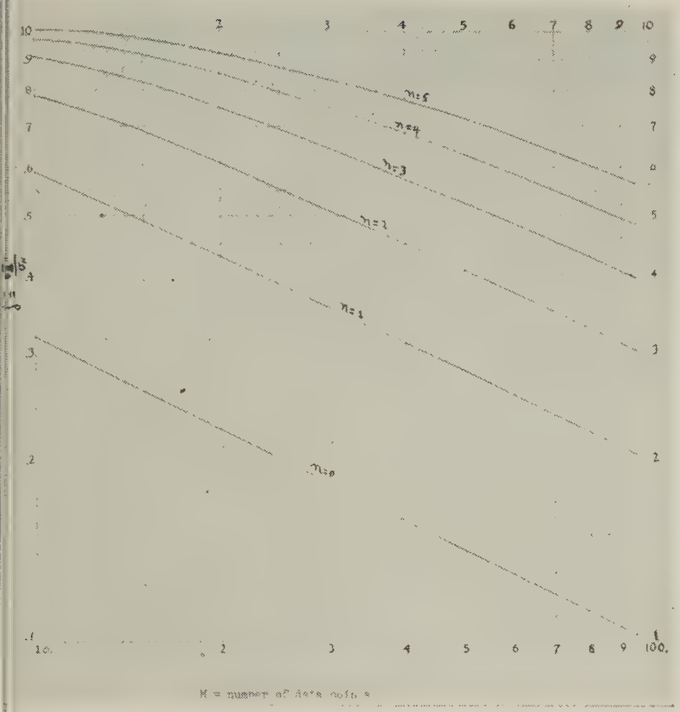
The functions satisfy the following recursion relationship

$$\xi_{v+1}(u) \equiv \xi_1(u)\xi_v(u) - \frac{v^2(M^2 - v^2)}{4(4v^2 - 1)}\xi_{v-1}(u). \tag{24}$$

As indicated the recursion is an identity so that by repeated differentiation one obtains

$$\xi_{v+1}^{(L)}(u) \equiv \xi_1(u)\xi_v^{(L)}(u) + L\xi_v^{(L-1)}(u) - \frac{v^2(M^2 - v^2)}{4(4v^2 - 1)}\xi_{v-1}^{(L)}(u), \tag{25}$$

[4] R. L. Anderson and E. E. Houseman, "Tables of Orthogonal Polynomials Values Extended to $N = 104$," Iowa State College of Agriculture and Mechanical Arts, Ames, Iowa, Res. Bull. 297; April, 1942.



ig. 1—Output of the digital filter is the least squares estimate of the input with zero lag.

$$\delta = \frac{\sigma_\Delta}{\sigma_N}$$

σ_Δ = rms error output (zero-lag estimation).
σ_N = rms error of noise.
n = degree of polynomial passed without error.
$(M-1)T$ = memory span of filter.

where

$$\xi_x^{(L)}(u) = 0, \quad L < 0, \quad x < 0, \quad \text{and} \quad L > x, \quad \xi_L^{(L)} = L!$$

so that

$$\xi_{v+1}^{(L)}(M+\alpha) = \left[\frac{2\alpha + M - 1}{2}\right]\xi_v^{(L)}(M+\alpha) + L\xi_v^{(L-1)}(M+\alpha) - \frac{v^2[M^2 - v^2]}{4(4v^2-1)}\xi_{v-1}^{(L)}(M+\alpha). \qquad (26)$$

Finally the sum of squares $S(L, M)$ is given by[5]

$$S(L, M) = \frac{(L!)^4 \prod_{j=-L}^{+L}(M-j)}{(2L)!(2L+1)!}. \qquad (27)$$

Higher order polynomials to degree 10 are listed by Allen.[5] A very complete table of the values of $\xi_v(u)$ for $v = 0$, to 5, u from $v+2$ to 104 is made available by Anderson and Houseman.[4]

Appendix II

Discussion of Asymptotic Formula

The asymptotic relation obtained by Johnson[1] is given by

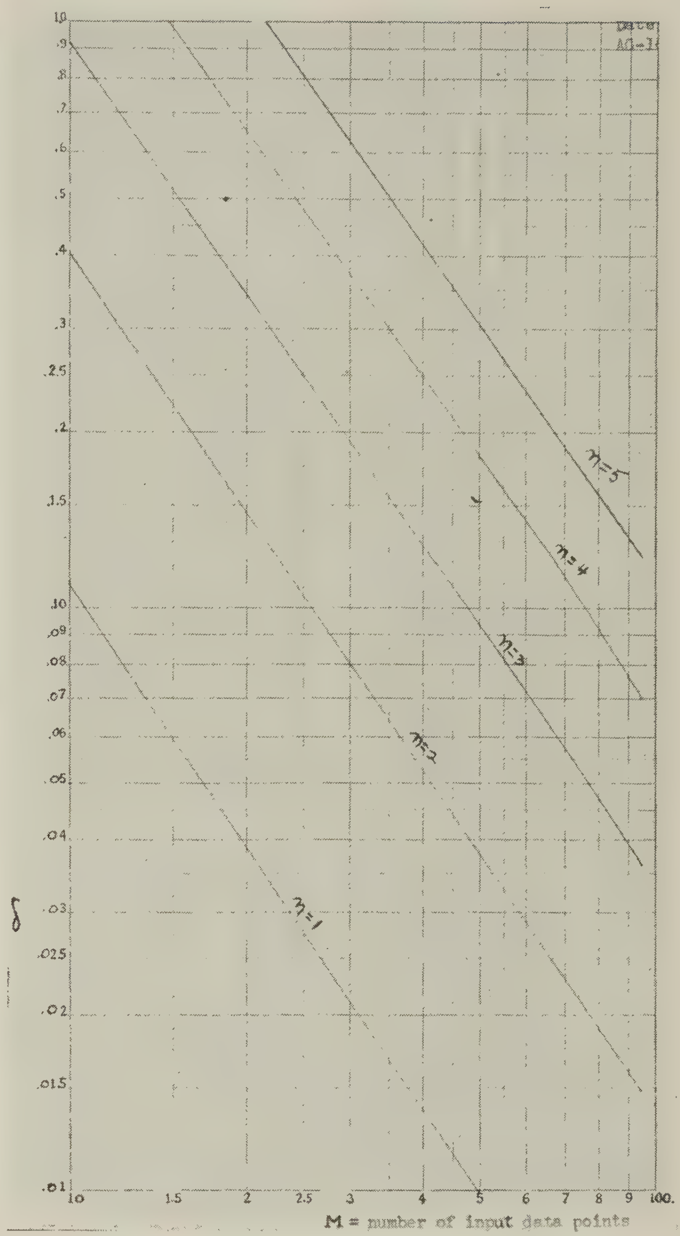


Fig. 2—Output of the digital filter is the least squares, zero-lag estimation of the first derivative of the input.

$$\delta = \frac{\sigma_\Delta T}{\sigma_N}$$

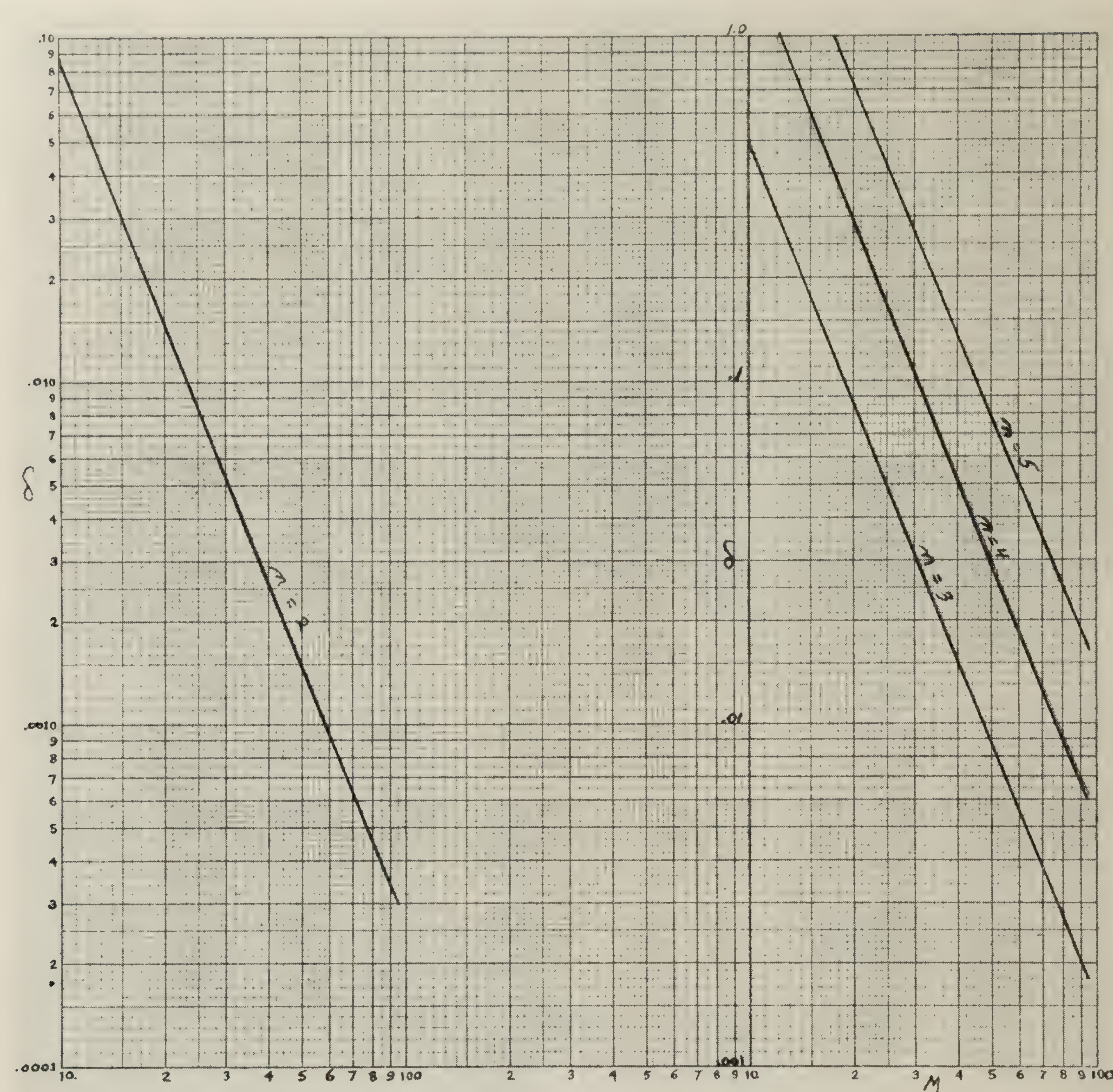
σ_Δ = rms of output.
σ_N = rms of input.
n = degree of polynomial passed by filter without error.
T = interval between samples.
$(M-1)T$ = finite memory of filter.

$$\delta \cong \frac{1}{(M-1)^{2K+1/2}} \frac{(K+n+1)!}{\sqrt{2K+1}(n-K)!K!}.$$

Eq. (22) presented in this report was obtained by using the curve fitting concepts and the orthogonal polynomials in the factorial form as presented in another paper.[6] The equations are asymptotically identical since the factor $(M-1) \to M$ for large M, and the remainder of the

[5] F. E. Allen, "The general form of the orthogonal polynomials for simple series with proofs of their simple properties," *Proc. Roy. Soc., Edinburgh*, vol. 50, pp. 310-320; 1935.

[6] W. E. Milne, "Numerical Calculus," Princeton Univ. Press, Princeton, N. J.; 1949.



M = number of input data points

Fig. 3—Output of the digital filter is the least squares, zero-lag ($\alpha = 0$) estimation of the second derivative of the input where

$$\delta = \frac{\sigma_\Delta T^2}{\sigma_N}$$

σ_Δ = rms of the output.
σ_N = rms of the noise.
n = degree of polynomial passed without error.
T = interval between samples.
$(M - 1)T$ = finite memory of filter.

equation give rise to the same coefficient. There are two properties of (22), however, which are apparent in the derivation.

First when the $n < K$, *e.g.*, the degree of the curve fitting polynomial is less than the order of the derivative, $\delta = 0$. To see this more clearly, consider a filter designed to pass linear functions ($n = 1$) and let the input be a constant or linear. Suppose the zero-lag estimate of the second derivative ($K = 2$) is the desired output. Thus, for any linear or constant input, the output should be zero. This result is assured if the weighting coefficients are all zero. The mean square error associated with the weighting coefficients is also zero, which is certainly the minimum. This result is implied if $(n - K)!$ is defined as infinity for $n < K$ but is not stated explicitly.

Second, (22) was obtained by expressing (17), as δ^2 is

in terms of the leading coefficient of the highest power of M (in both numerator and denominator) for each term. The factorial form of the orthogonal polynomials of Milne[6] leads directly to (22). As long as α is not a function of M, note that (22) is independent of α for $M \gg \alpha$. Thus to the degree of approximation represented by this form of the asymptotic formula, δ is independent of α. Eq. (22) with the factor $M^{-(2k+1)/2}$ was obtained directly by evaluating $\xi_{m\to v}^{(K)}$, *e.g.*, ($\alpha = 0$), and thus R of Tables I-III approaches zero as $M \to \infty$ faster than when the factor $(M - 1)^{(2k+1)/2}$ is used in the asymptotic formula.

Appendix III

"Best" Derivative Filters

In this Appendix the correspondence between the derivative of the least squares curve fit and the optimum derivative filter of another paper[2] will be shown. It is required to modify certain of those parameters to correspond to the notation of this paper.

The total number of samples is given by

$$M = m + 1. \tag{28}$$

Define $P_L(j\Delta t)$ of Blum's[2] (4) as

$$P_L(u) = \frac{\xi_L(u)}{[S(L, M)]^{1/2}}$$

$$u = 1, 2, \cdots, M \qquad L = 0, 1, 2, \cdots, n. \tag{29}$$

For the input model being considered, Blum's[2] (44) reduces to

$$|W| = P'|Q|, \tag{30}$$

since the matrix $V = \sigma^2 I$ (I is an identity matrix for uncorrelated noise), the matrix $\beta \equiv 0$ for $M(t) \equiv 0$, and the matrix $[PP'] = I$ because of the orthonormality of $P(u)$. (See Blum[2] and Fischer and Yates[3]).

The matrix P' is given by (25) of Blum[2] and becomes

$$P' = \begin{vmatrix} P_0(M) & P_1(M) & \cdots & P_n(M) \\ P_0(M-1) & P_1(M-1) & \cdots & P_n(M-1) \\ \vdots & & & \\ P_0(1) & P_1(1) & \cdots & P_n(1) \end{vmatrix}. \tag{31}$$

The matrix Q is given by (13) and (38) of Blum,[2]

$$Q_L = \frac{d^K}{dt^K} P_L(t) \bigg|_{t=M+\alpha} \equiv \frac{\xi_L^{(K)}(M+\alpha)}{[S(L, M)]^{1/2}}. \tag{32}$$

The intent in defining α is that $\alpha = 0$ represents a zero-lag estimate. The most current data point is measured at $t = M$, as opposed to $t = m\Delta t$ in the previous notation.[2]

On expending (30) for the vth weighting coefficient W_v one obtains

$$W_v = \sum_{L=0}^{n} P_L(M - v) Q_k(M + \alpha),$$

$$v = 0, 1, 2, \cdots, M - 1. \tag{33}$$

Substituting (29) and (32) one has

$$W_v = \sum_{L=K}^{n} \frac{\xi_L(M - v)\xi_L^{(K)}(M + \alpha)}{S(L, M)}. \tag{34}$$

Finally let $v = M - u$, $u = 1, 2, \cdots, M$, then

$$W_{M-u} = \sum_{L=K}^{n} \frac{\xi_L(u)\xi_L^{(K)}(M + \alpha)}{S(L, M)}, \tag{35}$$

which is identical to (11).

Derivative Estimates with Nonunity Sampling Intervals

The effect of $T \neq 1$ will be considered. The polynomials listed in Appendix I are nondimensional, thus when $T \neq 1$ one must substitute

$$\left[\frac{u - \bar{u}}{1}\right] = \left[\frac{u - \bar{u}}{T}\right]$$

$$u = T, 2T, \quad MT,$$

$$\bar{u} = \left[\frac{M+1}{2}\right]T.$$

An example of effect of taking derivatives is given as follows

$$\xi_2(u, T) = \left(\frac{u - \bar{u}}{T}\right)^2 - \frac{M^2 - 1}{2}$$

$$\frac{d\xi_2(u, T)}{du} = \frac{2}{T^2}(u - \bar{u}) = \frac{2}{T^2}\left(u - \left(\frac{M+1}{2}\right)T\right).$$

Let $u = vT$, then

$$\frac{d\xi_2(vT, T)}{du} = \frac{1}{T}\frac{d\xi_2}{du}(u, 1),$$

and

$$\frac{d^2\xi_2}{du^2}(vT, T) = \frac{1}{T^2}\frac{d^2\xi_2(u, 1)}{du^2}.$$

By similar methods one can show that

$$\frac{d^K\xi_L(vT, T)}{du^K} = \frac{1}{T^K}\frac{d^K\xi_L(u, 1)}{du^K}.$$

From (11) one sees that W_{m-u} is a linear combination of terms each of which are the Kth derivative with respect to u of $\xi_L(v\,T, T)|_{u=(M+\alpha)T}$ so that T^{-K} factors out of the expression for $W_{m-u,T}$ as stated in (16).

The Distribution of the Number of Crossings of a Gaussian Stochastic Process*

CARL W. HELSTROM†

Summary—It is shown how filtered Gaussian noise having a power spectrum which is a rational function of the square of the frequency can be represented as one component of a multidimensional Markov process. Methods are studied for obtaining the distribution of the number of times such a noise process crosses a given amplitude level in a fixed time interval. The generating function of this distribution is the solution of a Fokker-Planck type differential equation with appropriate boundary conditions. Integral equations are found for the generating function from which all the moments of the distribution can be calculated by iteration.

I. Introduction

A PROBLEM of interest in the theory of stochastic processes is to find the probability $p_n(t)$ that a random variable $x(t)$ crosses a given level, say $x = a$, n times in an interval of length t. For instance, this distribution determines the response to random noise of a device which measures frequency by counting the number of zero-crossings of its input. Also, the false-alarm probability for a gated radar detection system is $1 - p_0(t)$, where $p_0(t)$ is the probability that a given triggering level is not crossed by the noise in the gating interval of length t. The quantity $(-\partial p_0/\partial t)$ is the probability density function of the first time a noise wave crosses the level $x = a$, given certain initial conditions at $t = 0$. The problem of finding the latter has been solved for a process representing the Brownian motion of a free particle[1] and for one-dimensional Markov processes.[2] On the other hand, Rice[3] has given a formula for the mean number $\bar{n}$ of crossings in a given interval, and formulas for the variance of this number have been given by Steinberg *et al.*[4] and by Miller and Freund.[5]

In this paper we shall discuss the problem of finding the distribution $p_n(t)$ in the case of a stationary Gaussian random process $x(t)$ which is one component of a multidimensional Markov process. In the next section we shall show how any such Gaussian process having a power spectrum which is a rational function of ω^2, $\omega =$ (angular) frequency $= 2\pi f$, can be represented as one component of an n-dimensional Markov process, where ω^{2n} is the highest power of ω in the denominator of the power spectrum. The noise at the output of a linear filter made up of lumped circuit parameters is of this type when the input is white, Gaussian noise. In the last section we shall indicate that the generating function of the distribution $p_n(t)$ is the solution of a Fokker-Planck type differential equation with appropriate boundary conditions. Integral equations are obtained for the generating function which permit the determination of the moments of the distribution. Problems of this type can be attacked by the powerful general methods of Darling and Siegert,[6] of which our approach of Section III might be termed a discrete analog. However, our method seems to be more direct and more easily understandable in this particular case. Further details of our study of this problem are given in a research report.[7]

II. The Gaussian Markov Process

The principal example of the type of stochastic process to which the theory of the next section is applicable is that of filtered Gaussian noise. The random variable $x(t)$ is assumed for convenience to have zero mean, and its power spectrum is a rational function of ω^2; *i.e.*, it can be written

$$\Phi(\omega) = \frac{N(i\omega)N(-i\omega)}{P(i\omega)P(-i\omega)} = \left| \frac{N(i\omega)}{P(i\omega)} \right|^2 \tag{1}$$

where $N(z)$ and $P(z)$ are polynomials in z with real coefficients:

$$\begin{aligned} N(z) &= N_0 z^m + N_1 z^{m-1} + \cdots N_m \\ P(z) &= z^n + p_1 z^{n-1} + \cdots p_n \\ m &< n. \end{aligned} \tag{2}$$

The decomposition (1) is made so that the zeros of $N(z)$ and $P(z)$ have negative real parts. Such a process can be generated by passing white Gaussian noise $y(t)$ through a linear filter of admittance $Y(\omega) = N(i\omega)/P(i\omega)$; *i.e.*, $x(t)$ satisfies the differential equation

$$P(D)x(t) = N(D)y(t), \qquad D = d/dt. \tag{3}$$

We wish to find $(n - 1)$ other functions of t such that with $x(t)$ they form the n components of a Gaussian Markov process. The many ways of doing this can be derived from each other by linear transformations, and

* Manuscript received by the PGIT, March 25, 1957.

† Westinghouse Res. Labs., Pittsburgh 35, Pa.

[1] A. Blanc-Lapierre and R. Fortet, "Théorie des Fonctions Aléatoires," Masson et Cie, Paris, France; 1953.

[2] A. J. F. Siegert, "On the first passage time probability function," *Phys. Rev.*, vol. 81, pp. 617-623; February 15, 1951. References to the earlier literature are given here and in Blanc-Lapierre and Fortet, *loc. cit.*

[3] S. O. Rice, "The mathematical analysis of random noise," *Bell Sys. Tech. J.*, vol. 24, pp. 46-156; January, 1945.

[4] H. Steinberg, P. M. Schultheiss, C. A. Wogrin, and F. Zweig, "Short-time frequency measurement of narrow-band random signals by means of a zero counting process," *J. Appl. Phys.*, vol. 26, pp. 195-201; February, 1955.

[5] I. Miller and J. E. Freund, "Some results on the analysis of random signals by means of a cut-counting process," *J. Appl. Phys.*, vol. 27, pp. 1290-1293; November, 1956.

[6] D. A. Darling and A. J. F. Siegert, "A systematic approach to a class of problems in the theory of noise and other random phenomena—Part I," IRE Trans., vol. IT-3, pp. 32-37; March, 1957.

[7] C. W. Helstrom, "Level-Crossing Problems for Gaussian Stochastic Processes," Westinghouse Res. Labs., Pittsburgh, Pa., Rep. 8-1259-R5; March, 1957.

only one is presented here. The proofs of the following results are simple and are omitted for the sake of brevity; details can be found in another report.[7] The components $x_r(t)$ of the n-dimensional process are connected by the differential equations

$$dx_r/dt = x_{r+1} + C_r y(t), \qquad 0 \leq r < n-1,$$

$$dx_{n-1}/dt = -\sum_{s=1}^{n} p_s x_{n-s} + C_{n-1} y(t),$$

$$x_0 = x(t). \tag{4}$$

The constants C_r can be found from the contour integrals

$$C_r = \frac{1}{2\pi i} \int_C \frac{z^r N(z)}{P(z)}\, dz, \tag{5}$$

where C encloses all the poles of the integrand. To evaluate such an integral the transformation $z = 1/w$ is useful. Then it is easy to show that

$$C_r = 0, \qquad 0 \leq r \leq n-m-2 \; (m < n-1)$$
$$C_{n-m-1} = N_0, \tag{6}$$

while the rest of the C_r depend on the other coefficients of $N(z)$ and $P(z)$. Using (5) and eliminating the x_r successively from (4) one can show that these n-simultaneous first-order differential equations are equivalent to (3). Note that

$$x_r(t) = d^r x/dt^r, \qquad 0 \leq r \leq n-m-1 \tag{7}$$

so that the first $(n-m)$ components contain $x(t)$ and its first $(n-m-1)$ derivatives. The process $x(t)$ is said to be differentiable at most $(n-m-1)$ times.

In terms of the past history of $x(t)$ the values of all the $x_r(t)$ at time t can be found by means of the operators $0_r(D)$:

$$x_r(t) = 0_r(D)x(t), \qquad D = d/dt. \tag{8}$$

These operators are rational functions of D and can be found from the contour integral

$$0_r(p) = \frac{P(p)}{2\pi i N(p)} \int_C \frac{z^r N(z)}{(p-z)P(z)}\, dz, \tag{9}$$

where C encloses all poles of the integrand except that at $z = p$. In particular

$$0_r(p) = p^r, \qquad 0 \leq r \leq n-m-1,$$

$$0_{n-m}(p) = \frac{p^{n-m}N(p) - N_0 P(p)}{N(p)}. \tag{10}$$

In this operational notation the operator $(D-\mu)^{-1}$ is defined as usual by

$$(D-\mu)^{-1} f(t) = e^{\mu t} \int_{-\infty}^{t} e^{-\mu s} f(s)\, ds, \tag{11}$$

where the real part of μ is negative. $[N(D)]^{-1}$ is written as a product of such factors when it appears in the $0_r(D)$. The latter can always be evaluated in such a way that no more than $(n-m-1)$ differentiations of $x(t)$ are required. [Special consideration must be given to the case in which $N(z)$ vanishes for $z = 0$.[7]]

That the n functions $x_r(t)$ together form the components of a Gaussian Markov process is related to the fact that they are connected with each other and with $y(t)$ through the differential equations (4) of first order. If $y(t)$ is Gaussian, it can be shown that the joint conditional probability density function

$$p(\mathbf{x}, t|\mathbf{x}^0, t_0), \qquad \mathbf{x} = \{x_r(t)\}, \qquad \mathbf{x}^0 = \{x_r^0\} = \{x_r(t_0)\}$$

satisfies a Fokker-Planck equation of the form

$$\frac{1}{2} \sum_{r=n-m-1}^{n-1} \sum_{s=n-m-1}^{n-1} C_r C_s \frac{\partial^2 p}{\partial x_r \partial x_s} - \sum_{r=0}^{n-2} x_{r+1} \frac{\partial p}{\partial x_r} + \frac{\partial}{\partial x_{n-1}} \left(p \sum_{s=1}^{n} p_s x_{n-s} \right) - \frac{\partial p}{\partial t} = Lp - \partial p/\partial t$$
$$= -\prod_{r=0}^{n-1} \delta(x_r - x_r^0)\,\delta(t - t_0), \tag{12}$$

where L is a differential operator on the space coordinates $(x_0, x_1, \cdots x_{n-1})$. The coefficients in this equation were derived from (4) by the method discussed by Wang and Uhlenbeck.[8] By another method given there one can show that the solution of (12) which vanishes at infinity in all spatial directions, which is zero for $t < t_0$, and which approaches a product of delta functions as t approaches t_0 from above:

$$\lim_{t \to t_0} p(\mathbf{x}, t|\mathbf{x}^0, t_0) = \prod_{r=0}^{n-1} \delta(x_r - x_r^0), \qquad t > t_0, \tag{13}$$

is a Gaussian function of the form

$$p(\mathbf{x}, t|\mathbf{x}^0, t_0) = (2\pi)^{-n/2} \,|\det \phi_{rs}|^{-1/2} \cdot \exp\left[-\frac{1}{2} \sum_{r,s=0}^{n-1} \mu_{rs}(x_r - \bar{x}_r)(x_s - \bar{x}_s) \right], \tag{14}$$

where $\bar{x}_r = \overline{x_r(t)} = \mathbf{E}\{x_r, t \mid \mathbf{x}^0, t_0\}$ is the mean value of x_r at time t when the components had the given set of initial values $x_s^0 = x_s(t_0)$, $0 \leq r, s \leq n-1$. $\mathbf{E}$ = expectation value. The matrix $\|\mu_{rs}\|$ is the inverse of the cross-correlation matrix $\|\phi_{rs}\|$ of $x_r(t)$ and $x_s(t)$:

$$\phi_{rs}(t) = \mathbf{E}\{(x_r - \bar{x}_r)(x_s - \bar{x}_s), t|\mathbf{x}^0, t_0\}. \tag{15}$$

The $\overline{x_r(t)}$ and $\phi_{rs}(t)$ can be computed in terms of the spectrum (1) in a straightforward manner[7] by use of (8). The solution (14) is the basic transition probability density function for the n-dimensional Markov process.

As a function of the initial values $\mathbf{x}^0 = \{x_s^0\}$ the transition probability (14) satisfies the adjoint Fokker-Planck equation[1]

$$\frac{1}{2} \sum_{r=n-m-1}^{n-1} \sum_{s=n-m-1}^{n-1} C_r C_s \frac{\partial^2 p}{\partial x_r^0 \partial x_s^0} + \sum_{r=0}^{n-2} x_{r+1}^0 \frac{\partial p}{\partial x_r^0} - \left(\sum_{s=1}^{n} p_s x_{n-s}^0 \right) \frac{\partial p}{\partial x_{n-1}^0} + \frac{\partial p}{\partial t_0} = \tilde{L}_0 p + \partial p/\partial t_0$$
$$= -\prod_{r=0}^{n-1} \delta(x_r - x_r^0)\,\delta(t - t_0), \tag{16}$$

where $\tilde{L}_0$ is the differential operator adjoint to L; the subscript denotes that it operates on the $\{x_s^0\}$.

The Fokker-Planck equation (12) has the form of a conservation equation if it is written in terms of the n-dimensional divergence:

$$\operatorname{div} \mathbf{J} + \partial p/\partial t = 0, \qquad t > t_0, \tag{17}$$

where the "current" $\mathbf{J} = \{J_r\}$ is

$$J_r = x_{r+1}p - \frac{1}{2} C_r \sum_{s=n-m-1}^{n-1} C_s\, \partial p/\partial x_s,$$

$$0 \le r < n - 1$$

$$J_{n-1} = -p \sum_{s=1}^{n} p_s x_{n-s} - \frac{1}{2} C_{n-1} \sum_{s=n-m-1}^{n-1} C_s\, \partial p/\partial x_s. \tag{18}$$

Note that if $x(t)$ is differentiable at least once, $m < n - 1$, by (6) $C_0 = 0$ and $J_0 = x_1 p = \dot{x}p$, where the "velocity" $\dot{x} = dx/dt$. Then for the current in the x_0 direction we have only a transport term due to the velocity $\dot{x}$ and no gradient terms. Gradient terms disappear in all J_r for $0 \le r \le n - m - 2$.

It is convenient for purposes of visualization to speak of these equations as though they described the diffusion of particles having certain initial positions, velocities, etc., and we shall often do this in the sequel. This may be called the "Eulerian" approach, as contrasted to the "Lagrangian" method which considers the history of an individual particle. The results are applicable to other situations, of course, where $x(t)$ may be for instance a noise voltage. A thorough treatment of the two-dimensional case has been given by Wang and Uhlenbeck,[8] and they discuss the general problem of the thermal noise in a circuit containing N meshes of resistors, inductors, and capacitors, showing that this noise can be represented in terms of a Markov process of $2N$ dimensions.

III. The Distribution of the Number of Crossings

It is assumed that the random process $x(t)$ can be represented as one component of a Gaussian Markov process in n dimensions, $n > 1$, in the manner discussed above. It is further assumed that $x(t)$ is differentiable at least once ($m < n - 1$), so that we can define a velocity $\dot{x}(t) = x_1(t) = dx/dt$ as the second component of the Markov process. Let $p_n(t) = p_n(t \mid \mathbf{x}^0)$ be the probability that the random variable $x(t)$ starting at the initial value $\mathbf{x}^0 = (x_0, \dot{x}_0, x_{20}, x_{30}, \cdots x_{n-1,0})$ at $t = 0$ crosses the level $x = a$ n times in the interval $0 < t' < t$. In what follows we write this $p_n(t \mid x_0, \dot{x}_0)$ and omit the dependence on the variables $x_{r0} = x_r(0)$, $2 \le r \le n - 1$, but the presence of this dependence must be kept in mind. In expressions involving integrals, integrations are also taken over the entire ranges of these $(n - 2)$ unwritten variables. We shall find equations for the generating function of this distribution:

$$h(t|x_0, \dot{x}_0; z) = \sum_{n=0}^{\infty} z^n p_n(t|x_0, \dot{x}_0). \tag{19}$$

If $x_0 < a$ we denote this by $h^-(t \mid x_0, \dot{x}_0; z)$, if $x_0 > a$ by $h^+(t \mid x_0, \dot{x}_0; z)$. If some distribution of initial values at $t = 0$ is given, $h(t \mid x_0, \dot{x}_0; z)$ can be multiplied by it at the end and the product integrated to determine the generating function for the given problem. Thus we are dealing here with the generating function of a conditional distribution.

The quantity $p_0(t \mid x_0, \dot{x}_0) = h(t \mid x_0, \dot{x}_0; 0)$ is the probability that $x(t)$ does not cross $x = a$ in the interval $0 < t' < t$. For the Markov processes they considered, it has been pointed out by Siegert[2] and Wasow[9] that $p_0(t \mid x_0, \dot{x}_0)$ satisfies a partial differential equation involving the adjoint Fokker-Planck operator $\tilde{L}$ of (16):

$$\tilde{L}_0 p_0(t|x_0, \dot{x}_0) - \partial p_0/\partial t = 0, \qquad t > 0. \tag{20}$$

For the n-dimensional process of Section II, the boundary conditions on this equation are,[7] for $x_0 < a$,

$$p_0(t|a, \dot{x}_0) = 0, \qquad \dot{x}_0 > 0$$

$$p_0(0|x_0, \dot{x}_0) = 1, \qquad x_0 < a. \tag{21}$$

Further we can define a function $G_-(x, \dot{x}, t \mid x_0, \dot{x}_0, t_0)$ as the joint density function of $x, \dot{x}, x_2, \cdots x_{n-1}$ at time t, given that the trajectories started at $x_0, \dot{x}_0$, etc. at time t_0 with $x_0 < a$, and remained in the region $x < a$ for the entire interval $t_0 < t' < t$. Then

$$p_0(t|x_0, \dot{x}_0) = \int_{-\infty}^{a} dx \int_{-\infty}^{\infty} d\dot{x} G_-(x, \dot{x}, t|x_0, \dot{x}_0, 0). \tag{22}$$

The function $G_-(x, \dot{x}, t \mid x_0, \dot{x}_0, t_0)$ satisfies a Fokker-Planck equation of the form (12) as a function of the final coordinates $x, \dot{x}, \cdots t$, with the additional boundary condition[8] that it vanishes at $x = a$ for $\dot{x} < 0$. As a function of the initial coordinates $x_0, \dot{x}_0, \cdots t_0$ it satisfies (16) with the additional boundary condition[7] that it vanishes at $x_0 = a$ for $\dot{x}_0 > 0$. A similar function $G_+(x, \dot{x}, t \mid x_0, \dot{x}_0, t_0)$ can be defined for trajectories which remain in the region $x > a$ for the whole interval $t_0 < t' < t$.

Referring to (19) we consider a random process in which the trajectories $x(t)$ are the same as for free diffusion; that is, their development is described by the Fokker-Planck equation (12). Each time a trajectory (or "diffusing particle") crosses $x = a$, there are probabilities $z < 1$ that it will survive and $(1 - z)$ that it will be removed from the system; but if it survives it continues on its way with no change in velocity or in any other coordinate. Then $h(t \mid x_0, \dot{x}_0; z)$ is the fraction of such particles remaining at time t, out of all those starting at $(x_0, \dot{x}_0, \cdots)$ at time $t = 0$.

It is plausible $h(t \mid x_0, \dot{x}_0; z)$ is a solution of an adjoint Fokker-Planck equation of the same type as (20), and a detailed consideration shows this is so.[7] That is,

$$\tilde{L}_0 h(t|x_0, \dot{x}_0; z) - \partial h/\partial t = 0, \qquad t > 0. \tag{23}$$

To find the boundary conditions, consider one particle starting just below $x = a$ with positive velocity, and

[8] M. C. Wang and G. E. Uhlenbeck, "On the theory of the Brownian motion II," *Rev. Mod. Phys.*, vol. 17, pp. 323-342; April, July, 1945.

[9] W. Wasow, "On the duration of random walks," *Ann. Math. Stat.*, vol. 22, pp. 199-216; July, 1951.

another starting just above $x = a$ with the same initial conditions. The former will have one more crossing in $0 < t' < t$ than the latter. Denoting the respective probabilities by the superscripts $-$ and $+$,

$$p_n^-(t) = p_{n-1}^+(t), \qquad n > 0,$$
$$p_0^-(t) = 0. \tag{24}$$

Hence from (19)

$$h^-(t; z) = \sum_{n=0}^{\infty} z^n p_n^-(t) = \sum_{n=1}^{\infty} z^n p_{n-1}^+(t)$$
$$= zh^+(t; z), \qquad \dot{x}_0 > 0. \tag{25}$$

Using a similar argument for $\dot{x}_0 < 0$ we get the boundary conditions

$$h^-(t|a, \dot{x}_0; z) = zh^+(t|a, \dot{x}_0; z), \qquad \dot{x}_0 > 0$$
$$h^+(t|a, \dot{x}_0; z) = zh^-(t|a, \dot{x}_0; z), \qquad \dot{x}_0 < 0 \tag{26}$$

where $h^-(t \mid a, \dot{x}_0; z)$ is the limit of $h^-(t \mid x_0, \dot{x}_0; z)$ as $x_0 \to a$ from below, while in $h^+(t|a, \dot{x}_0; z)$ $x_0 \to a$ from above. We also have the condition

$$h(0|x_0, \dot{x}_0; z) = 1, \tag{27}$$

since $p_0(0 \mid x_0, \dot{x}_0) = 1$, $p_n(0 \mid x_0, \dot{x}_0) = 0$, $n > 0$. Here we have a discrete analog of the method of Darling and Siegert;[6] a crossing of the boundary $x = a$ corresponds to a "collision" in the language of their article.

We now write down integral equations for the generating function which can be obtained[7] by applying Green's theorem[10] to (23). In the first set the kernels are the functions G_- and G_+ defined above:

$$h^-(t|x_0, \dot{x}_0; z) = h^-(t|x_0, \dot{x}_0; 0) + z \int_0^t dt_1 \int_0^\infty \dot{x}_1 h^+(t - t_1|a, \dot{x}_1; z) \cdot G_-(a, \dot{x}_1, t_1|x_0, \dot{x}_0, 0)\, d\dot{x}_1, \qquad x_0 < a, \tag{28}$$

$$h^+(t|x_0, \dot{x}_0; z) = h^+(t|x_0, \dot{x}_0; 0) + z \int_0^t dt_1 \int_{-\infty}^0 |\dot{x}_1|\, h^-(t - t_1|a, \dot{x}_1; z) \cdot G_+(a, \dot{x}_1, t_1|x_0, \dot{x}_0, 0)\, d\dot{x}_1, \qquad x_0 > a. \tag{29}$$

In the application of Green's theorem, one obtains a term depending on the component of the current **J** normal to the surface $x = a$, with an integration over that surface and over the time interval $0 < t' < t$. Because $x(t)$ is differentiable, $m < n - 1$, the component of **J** in the x-direction consists only of a transport term and no gradient terms [see (18) *et seq.*], and one obtains the second term on the right of (28) and (29), after using (26). It is because a velocity is definable for these differentiable processes $x(t)$ that the integral equations possess a comparatively simple form.

The meaning of (28) is clear: the first term on the right is the number of particles at time t that have never crossed $x = a$, while the second term classifies the remaining particles according to the first time they crossed $x = a$ with positive velocity. On this crossing only a fraction z survived, and each of these crossing with velocity $\dot{x}_1 > 0$ in t_1 to $t_1 + dt_1$ had a probability $h^+(t - t_1) \mid a, \dot{x}_1; z)$ of lasting until time t. A similar explanation holds for (29). The distribution $p_n(t \mid x_0, \dot{x}_0)$ can be found by iteration of (28) and (29). For example, for $x_0 < a$,

$$p_1(t) = \int_0^t dt_1 \int_0^\infty \dot{x}_1 h^+(t - t_1|a, \dot{x}_1; 0) \cdot G_-(a, \dot{x}_1, t_1|x_0, \dot{x}_0, 0)\, d\dot{x}_1,$$

$$p_2(t) = \int_0^t dt_1 \int_0^{t_1} dt_2 \int_0^\infty \dot{x}_1\, d\dot{x}_1 \int_{-\infty}^0 |\dot{x}_2|\, d\dot{x}_2 \cdot h^-(t - t_2|a, \dot{x}_2; 0) G_+(a, \dot{x}_2, t_2|a, \dot{x}_1, t_1) \cdot G_-(a, \dot{x}_1, t_1|x_0, \dot{x}_0, 0), \tag{30}$$

results which can of course be written down from first principles.

Since the density functions in (28) and (29) are usually not known, it is useful to find integral equations involving the known transition probability function (14). Application of Green's theorem yields[7]

$$h^-(t|x_0, \dot{x}_0; z) = \int_{-\infty}^a dx \int_{-\infty}^\infty d\dot{x}\, p(x, \dot{x}, t|x_0, \dot{x}_0, 0) + z \int_0^t dt_1 \int_0^\infty \dot{x}_1 h^+(t - t_1\, a, \dot{x}_1; z) \cdot p(a, \dot{x}_1, t_1|x_0, \dot{x}_0, 0)\, d\dot{x}_1 + \int_0^t dt_1 \int_{-\infty}^0 \dot{x}_1 h^-(t - t_1|a, \dot{x}_1; z) \cdot p(a, \dot{x}_1, t_1|x_0, \dot{x}_0, 0)\, d\dot{x}_1, \qquad x_0 < a \tag{31}$$

$$0 = \int_a^\infty dx \int_{-\infty}^\infty d\dot{x}\, p(x, \dot{x}, t|x_0, \dot{x}_0, 0) - \int_0^t dt_1 \int_0^\infty \dot{x}_1 h^+(t - t_1|a, \dot{x}_1; z) \cdot p(a, \dot{x}_1, t_1|x_0, \dot{x}_0, 0)\, d\dot{x}_1 - z \int_0^t dt_1 \int_{-\infty}^0 \dot{x}_1 h^-(t - t_1|a, \dot{x}_1; z) \cdot p(a, \dot{x}_1, t_1|x_0, \dot{x}_0, 0)\, d\dot{x}_1, \qquad x_0 < a \tag{32}$$

$$h^+(t|x_0, \dot{x}_0; z) = \int_a^\infty dx \int_{-\infty}^\infty d\dot{x}\, p(x, \dot{x}, t|x_0, \dot{x}_0, 0) - \int_0^t dt_1 \int_0^\infty \dot{x}_1 h^+(t - t_1|a, \dot{x}_1; z) \cdot p(a, \dot{x}_1, t_1|x_0, \dot{x}_0, 0)\, d\dot{x}_1 - z \int_0^t dt_1 \int_{-\infty}^0 \dot{x}_1 h^-(t - t_1|a, \dot{x}_1; z) \cdot p(a, \dot{x}_1, t_1|x_0, \dot{x}_0, 0)\, d\dot{x}_1, \qquad x_0 > a \tag{33}$$

[10] P. M. Morse and H. Feshbach, "Methods of Theoretical Physics," McGraw-Hill Book Co., Inc., New York, N. Y.; 1953. See Sections 7.1 and 7.5.

$$0 = \int_{-\infty}^{a} dx \int_{-\infty}^{\infty} d\dot{x}\, p(x, \dot{x}, t|x_0, \dot{x}_0, 0)$$
$$+ z \int_0^t dt_1 \int_0^{\infty} \dot{x}_1 h^+(t - t_1|a, \dot{x}_1; z)$$
$$\cdot p(a, \dot{x}_1, t_1|x_0, \dot{x}_0, 0)\, d\dot{x}_1$$
$$+ \int_0^t dt_1 \int_{-\infty}^{0} \dot{x}_1 h^-(t - t_1|a, \dot{x}_1; z)$$
$$\cdot p(a, \dot{x}_1, t_1|x_0, \dot{x}_0, 0)\, d\dot{x}_1, \qquad x_0 > a. \tag{34}$$

For $z = 1$ these equations reduce to various forms of the conservation law (17), since

$$h(t|x_0, \dot{x}_0; 1) = 1; \tag{35}$$

in this case there is no loss of particles. The meaning of the equations is also clear when $z = 0$, $h(t \mid x_0, \dot{x}_0; 0) = p_0(t \mid x_0, \dot{x}_0)$; for instance (31) then shows that the number of particles in $x < a$ at time t (the first term on the right) is made up of those which have never crossed $x = a$ (the term on the left) plus those which have crossed $x = a$ at least once, these being classified according to the last time they crossed $x = a$ (the negative of the third term on the right). For $0 < z < 1$ the accounting of the numbers of particles is more complicated.

The most symmetrical form of these equations is obtained by adding (31) to (32), and (33) to (34). Since the integral of $p(x, \dot{x}, t \mid x_0, \dot{x}_0, 0)$ over all space coordinates $x, \dot{x}, \cdots x_{n-1}$ must be unity, we get

$$h^-(t|x_0, \dot{x}_0; z) = 1 + (z - 1) \int_0^t dt_1$$
$$\int_0^{\infty} \dot{x}_1 h^+(t - t_1|a, \dot{x}_1; z) p(a, \dot{x}_1, t_1|x_0, \dot{x}_0, 0)\, d\dot{x}_1$$
$$+ (z - 1) \int_0^t dt_1 \int_{-\infty}^{0} |\, \dot{x}_1 \,|\, h^-(t - t_1|a, \dot{x}_1; z)$$
$$\cdot p(a, \dot{x}_1, t_1|x_0, \dot{x}_0, 0)\, d\dot{x}_1, \qquad x_0 < a \tag{36}$$

$$h^+(t|x_0, \dot{x}_0; z) = 1 + (z - 1) \int_0^t dt_1$$
$$\cdot \int_0^{\infty} \dot{x}_1 h^+(t - t_1|a, \dot{x}_1; z) p(a, \dot{x}_1, t_1|x_0, \dot{x}_0, 0)\, d\dot{x}_1$$
$$+ (z - 1) \int_0^t dt_1 \int_{-\infty}^{0} |\, \dot{x}_1 \,|\, h^-(t - t_1|a, \dot{x}_1; z)$$
$$\cdot p(a, \dot{x}_1, t_1|x_0, \dot{x}_0, 0)\, d\dot{x}_1, \qquad x_0 > a. \tag{37}$$

Other integral equations can be obtained[7] in terms of a function $H(x, \dot{x}, t \mid x_0, \dot{x}_0, 0; z)$ which describes the density distribution of particles in our artificial absorption process.

By (19) the moments of the distribution $p_n(t \mid x_0, \dot{x}_0)$ are related to the coefficients of a power-series expansion of $h(t \mid x_0, \dot{x}_0; z)$ about the point $z = 1$. This expansion is obtained immediately by solving (36) and (37) by iteration:

$$h^-(t|x_0, \dot{x}_0; z) = 1 + (z - 1) \int_0^t dt_1$$
$$\cdot \int_{-\infty}^{\infty} |\, \dot{x}_1 \,|\, p(a, \dot{x}_1, t_1|x_0, \dot{x}_0, 0)\, d\dot{x}_1$$
$$+ \lim_{\epsilon \to 0} (z - 1)^2 \int_0^t dt_2 \int_0^{t_2} dt_1 \int_{-\infty}^{\infty} |\, \dot{x}_2 \,|\, d\dot{x}_2 \int_0^{\infty} \dot{x}_1\, d\dot{x}_1$$
$$\cdot p(a - \epsilon, \dot{x}_2, t_2|a, \dot{x}_1, t_1) p(a, \dot{x}_1, t_1|x_0, \dot{x}_0, 0)$$
$$+ \lim_{\epsilon \to 0} (z - 1)^2 \int_0^t dt_2 \int_0^{t_2} dt_1 \int_{-\infty}^{\infty} |\, \dot{x}_2 \,|\, d\dot{x}_2 \int_{-\infty}^{0} |\, \dot{x}_1 \,|\, d\dot{x}_1$$
$$\cdot p(a + \epsilon, \dot{x}_2, t_2|a, \dot{x}_1, t_1) p(a, \dot{x}_1, t_1|x_0, \dot{x}_0, 0)$$
$$+ \cdots, x_0 < a, \qquad \epsilon > 0. \tag{38}$$

Thus the second term yields Rice's result[3] for the mean number $\bar{n}$ of crossings in an interval of length t:

$$\bar{n}(t|x_0, \dot{x}_0) = \int_0^t dt_1 \int_{-\infty}^{\infty} |\, \dot{x}_1 \,|\, p(a, \dot{x}_1, t_1|x_0, \dot{x}_0, 0)\, d\dot{x}_1. \tag{39}$$

In most problems the particles are initially distributed in accordance with the equilibrium distribution

$$W(x, \dot{x}) = \lim_{t \to \infty} p(x, \dot{x}, t|x_0, \dot{x}_0, t_0).$$

Multiplying (39) by $W(x_0, \dot{x}_0)$ and integrating over the entire ranges of these variables, we obtain the usual formula for the average number of crossings:

$$\bar{n}(t) = t \int_{-\infty}^{\infty} |\, \dot{x} \,|\, W(a, \dot{x})\, d\dot{x}. \tag{40}$$

This formula has been applied by Middleton[11] to compute the false-alarm rate of a triggered circuit such as that in a radar detection system.

From (19) and (38) we see that the mean-square of the number of crossings, $\overline{n^2}$, is given by

$$\overline{n^2} = \bar{n} + 2 \lim_{\epsilon \to 0} \int_0^t dt_2 \int_0^{t_2} dt_1 \int_{-\infty}^{\infty} |\, \dot{x}_2 \,|\, d\dot{x}_2$$
$$\left[\int_0^{\infty} \dot{x}_1\, d\dot{x}_1 p(a - \epsilon, \dot{x}_2, t_2; a, \dot{x}_1, t_1|x_0, \dot{x}_0, 0) \right.$$
$$\left. + \int_{-\infty}^{0} |\, \dot{x}_1 \,|\, d\dot{x}_1 p(a + \epsilon, \dot{x}_2, t_2; a, \dot{x}_1, t_1|x_0, \dot{x}_0, 0) \right],$$
$$x_0 < a, \qquad \epsilon > 0, \tag{41}$$

where we have integrated over the coordinates corresponding to $x_r(t)$, $2 \leq r \leq n - 1$, expressing our result in terms of the joint conditional probability density function $p(x, \dot{x}, t; x_1, \dot{x}_1, t_1 \mid x_0, \dot{x}_0, 0)$ of $x, \dot{x}$ at time t and of $x_1, \dot{x}_1$ at time t_1, given the initial values $x_0, \dot{x}_0$ at time $t = 0$. Eq. (41) corresponds to the result of Steinberg *et al*,[4] and of Miller and Freund,[5] although these authors use different methods of avoiding the singularity at $t_1 = t_2$ in the integrand. Moments of higher order can be obtained by continuing the series solution (38).

We remark in closing that these methods can be applied

[11] D. Middleton, "Spurious signals caused by noise in triggered circuits," *J. Appl. Phys.*, vol. 19, pp. 817-830; September, 1948.

to determine the distributions of the number of maxima or minima of a random function in a given interval. The relevant boundary in n-dimensional space is now the hyperplane $\dot{x} = 0$. If the process is twice differentiable $(m < n - 2)$, we can define an acceleration variable $\ddot{x}(t) = x_2(t) = d^2x/dt^2$ which is also a component of the Markov process. Maxima then correspond to the trajectory in n-dimensional space crossing $\dot{x} = 0$ in $\ddot{x} < 0$, while minima correspond to crossing it in $\ddot{x} > 0$. The methods of this section will yield, *e.g.*, in place of (40), Rice's expression[3] for the average number of maxima in a range x to $x + dx$ in an interval of length t. Such quantities as the variance of the total number of maxima and minima can be found by similarly extending the analysis.

Unfortunately none of the methods of this section is simple for computation. To solve the Fokker-Planck differential equations would no doubt require a high-speed computing machine. The integral equations are made difficult to handle by the singularity of the kernels, and the Wiener-Rice series[3] which result when they are solved by iteration contain cumbersome integrals even in terms of low order. To attain results of practical importance, approximations must be made. It is hoped that the theory outlined here will be of use in such an effort.

An Analysis of Coherent Integration and Its Application to Signal Detection*

K. S. MILLER† AND R. I. BERNSTEIN‡

Summary—An important characteristic of coherent integrators is that their effective bandwidth decreases as the integration time increases. If it is only known that a weak signal occurs somewhere in a given frequency range, then the number of integration channels required to cover the specified range increases as the amount of coherent integration is increased. However, each integration channel can independently cause a false alarm, although only the particular channel in which the signal appears can cause a true alarm. The question arises therefore whether it is profitable to lengthen the coherent integration period to increase the signal-to-noise ratio when doing so requires an increase in the number of integration channels. This problem is investigated analytically. Numerical results appropriate for system design are presented as a series of graphs of missed-signal probability vs number of integration channels, with initial signal-to-noise ratio and over-all false alarm probability as parameters.

Also included is a detailed analysis of statistical properties of ideal and approximate ideal coherent integrators.

I. Introduction

MANY radar systems of current interest employ coherent integration for the detection and identification of weak signals in Gaussian noise. In the case of coherent cw radars the echo, after coherent heterodyning, is a sine wave whose existence indicates the presence of a target and whose frequency reveals information concerning the target. In the case of coherent pulse radars the echo, after coherent heterodyning, is a pulse train with a sinusoidal envelope. We shall assume that the pulse envelope is converted to cw by subsequent processing or otherwise placed on a basis which allows it to be considered in the same manner as a continuous sine wave. A common method of converting the amplitude modulated pulse train to a continuous wave employs a pulse stretcher, or "boxcar," circuit followed by a filter which compensates for the pulse stretcher's spectral characteristic.

Properties of coherent integrators will be defined precisely and investigated in Section II. The application here to signal detection will be the problem of devising a procedure to test the hypothesis that a sinusoidal signal exists in the presence of Gaussian noise. (It is desirable also to identify the frequency of the signal.) Towards this end the enhancement of the signal-to-noise ratio achieved by coherent integration will be analyzed in Section III.

The quality, or strength, of a statistical hypothesis test is specified by giving the false alarm and missed-signal probabilities. A false alarm, known in statistical terminology as a Type I error, occurs if it is concluded that a signal is present when no signal really exists. A missed signal, or Type II error, occurs when a signal really exists and it is concluded that none is present. The missed-signal probability is one minus the detection probability. The hypothesis test can be implemented by applying the combination of signal and noise to a threshold device which indicates whether a preset amplitude level is exceeded. If the threshold is exceeded, the hypothesis is accepted that the signal has occurred. A knowledge of the amplitude probability density functions of the noise

* Manuscript received by the PGIT, June 25, 1957.

† Dept. of Mathematics, New York University, New York 53, N. Y., and Columbia University Electronics Res. Labs., New York 27, N. Y.

‡ Columbia University Electronics Res. Labs., New York 27, N. Y.

alone and of the signal plus noise is required in order to ascertain the false alarm and missed-signal probabilities that result from the choice of an acceptance threshold voltage. This problem of detection and its ramifications will be discussed in Section IV.

II. Coherent Integration

In order to improve the quality of the signal detection test, we desire to improve the signal-to-noise ratio. It will be shown that this enhancement can be achieved by the use of coherent integration. This section will be devoted accordingly to a discussion of coherent integrators.

In Part A the ideal coherent integrator (ICI) will be defined and some of its features discussed. In Part B we shall investigate the statistical properties of inputs and outputs, and in Part C we shall compute the cumulative distribution of the envelopes of the outputs of a bank of n ICI. A device which performs almost as an ICI will be described in Part D. Finally, in part E we shall analyze some of the statistical properties of inputs and outputs of this physical device.

A. The Ideal Coherent Integrator

An ideal coherent integrator essentially integrates the quadrature components of the envelope of an applied sinusoid. We make this notion precise. Let $e_1(t)$ be a given signal. Then with respect to the frequency ω_0, $e_1(t)$ may be written as

$$e_1(t) = a_1(t) \cos \omega_0 t + b_1(t) \sin \omega_0 t. \tag{1}$$

The above representation is not unique. For example, $e_1(t)$ could also be written as $e_1(t) = c_1(t) \cos \omega_0 t + d_1(t) \sin \omega_0 t$ where $c_1(t) = a_1(t) + b_1(t) \tan \omega_0 t$ and $d_1(t) = 0$. However, there is only one expression of the form of (1) where $[a_1^2(t) + b_1^2(t)]^{1/2}$ is the *envelope*[1] of the waveform $e_1(t)$ with respect to the frequency ω_0. It will be assumed in future work that (1) represents this unique expansion of $e_1(t)$ in terms of two amplitude-modulated sinusoids in quadrature at the frequency ω_0. An ICI tuned to the frequency ω_0 is defined as a device whose response $e_2(t)$ to the applied input $e_1(t)$ is

$$e_2(t) = \cos \omega_0 t \int_0^t a_1(\xi)\, d\xi + \sin \omega_0 t \int_0^t b_1(\xi)\, d\xi. \tag{2}$$

To illustrate this point Fig. 1 shows the response of an ICI to a constant amplitude sine wave whose frequency is the same as the integrator's tuned frequency. Thus, $e_1(t) = \sin \omega_0 t$ and $e_2(t) = t \sin \omega_0 t$. We shall call ω_0 the "tuned frequency" of the coherent integrator. The integrator is "ideal" if the coefficients $a_2(t)$ and $b_2(t)$ of $e_2(t)$ are exactly the integrals of $a_1(t)$ and $b_1(t)$, respectively. In many practical cases these integrals are only approximated,

[1] In general, if $y = f(t, \omega)$ where ω is a parameter, the envelope may be determined by eliminating ω between the equations $y - f = 0$ and $\partial f/\partial \omega = 0$.

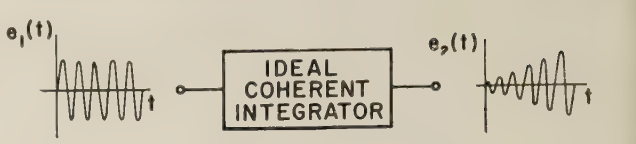


Fig. 1—Output $e_2(t)$ of ideal coherent integrator with input $e_1(t)$ a constant amplitude sinusoid at tuned frequency of integrator.

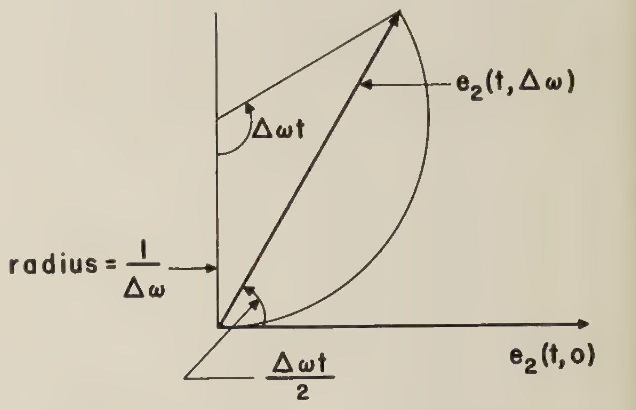


Fig. 2—Phasor representation of coherent integrator response.

and then the device used as a coherent integrator is not ideal.

An interesting phasor interpretation of the response of an ideal coherent integrator to an input signal of frequency unequal to the tuned frequency of the ICI may be given. While this result is not explicitly used in this paper, it has been found to be a useful tool in certain associated work.

In order to develop this phasor representation, suppose that

$$e_1(t, \Delta\omega) = \cos (\omega_0 + \Delta\omega) t$$

is a constant amplitude sinusoid at a frequency displaced by an amount $\Delta\omega$ from the integrator's tuned frequency. Then the input to the ICI will be

$$e_1(t, \Delta\omega) = \cos \Delta\omega t \cos \omega_0 t - \sin \Delta\omega t \sin \omega_0 t.$$

From (2) it follows that

$$e_2(t, \Delta\omega) = \cos \omega_0 t \int_0^t \cos \Delta\omega\xi\, d\xi - \sin \omega_0 t \int_0^t \sin \Delta\omega\xi\, d\xi$$

which may be written as

$$e_2(t, \Delta\omega) = \frac{\sin \frac{\Delta\omega t}{2}}{\frac{\Delta\omega}{2}} \cos \left(\omega_0 t + \frac{\Delta\omega t}{2}\right). \tag{3}$$

If $\Delta\omega = 0$, of course, $e_1(t, \Delta\omega) = e_1(t, 0) = \cos \omega_0 t$ and the output of the ICI is $e_2(t, 0) = t \cos \omega_0 t$.

Fig. 2 is a phasor diagram of these results portrayed on a set of axes which revolve at the frequency ω_0. The tip of the vector $e_2(t, 0)$ moves at a uniform rate along the horizontal axis and has a length t. The tip of vector $e_2(t, \Delta\omega)$ moves at a constant angular rate on the circumference of a circle with center at $(0, 1/\Delta\omega)$ and radius $1/\Delta\omega$. Although the developed length of the trajectory

of $e_2(t, \Delta\omega)$ is always the same as the length of $e_2(t, 0)$, the vector itself experiences periodic nulls.

We wish to draw one further conclusion concerning this phasor representation. Let $E(t, \Delta\omega)$ be the ratio of the phasors representing $e_2(t, \Delta\omega)$ and $e_2(t, 0)$. That is,

$$E(t, \Delta\omega) = \frac{e_2(t, \Delta\omega)}{e_2(t, 0)} = \frac{\dfrac{\sin\frac{\Delta\omega t}{2}}{\frac{\Delta\omega}{2}} \Big| \frac{\Delta\omega t}{2}}{t \,\underline{|0}} = \frac{\sin\frac{\Delta\omega t}{2}}{\frac{\Delta\omega t}{2}} \Big| \frac{\Delta\omega t}{2}. \quad (4)$$

Among other things, this result shows that the effective bandwidth of an ICI decreases as the integration time t increases. This relation is exploited in Section IV, where we discuss the multichannel problem.

B. Statistical Properties of Ideal Coherent Integrators

Suppose that the input to an ICI tuned to the frequency ω_0 is statistically stationary Gaussian noise, $i(t)$. We shall assume that $i(t)$ has mean zero and correlation function $\psi(\tau)$. Rice[2] has shown that $i(t)$ may be written as

$$i(t) = i_C(t) \cos \omega_0 t + i_S(t) \sin \omega_0 t$$

where i_C and i_S are independent Gaussian variates. It follows immediately that

$$\langle i \rangle = \langle i_C \rangle = \langle i_S \rangle = 0$$

(where $\langle \cdots \rangle$ indicates ensemble average) and

$$\langle i^2 \rangle = \langle i_C^2 \rangle = \langle i_S^2 \rangle = \psi_0$$

where $\psi_0 \equiv \psi(0)$.

The output of the ICI is, by definition,

$$I(t) = \left[\int_0^t i_C(\xi)\, d\xi\right] \cos \omega_0 t + \left[\int_0^t i_S(\xi)\, d\xi\right] \sin \omega_0 t.$$

The stochastic integrals $I_C(t) = \int_0^t i_C(\xi) d\xi$ and $I_S(t) = \int_0^t i_S(\xi) d\xi$ are also independent Gaussian variates. Their means are zero since

$$\langle I_C(t) \rangle = \int_0^t \langle i_C(\xi) \rangle\, d\xi = 0$$

with a similar formula for $\langle I_S \rangle$. Of course, $\langle I \rangle$ is also zero. However, their variances are time dependent. To compute the variance of $I(t)$ write

$$\langle I^2(t) \rangle = \langle I_C^2(t) \rangle \cos^2 \omega_0 t + \langle I_S^2(t) \rangle \sin^2 \omega_0 t + 2\langle I_C(t) I_S(t) \rangle \cos \omega_0 t \sin \omega_0 t.$$

But

$$\langle I_C^2(t) \rangle = \int_0^t \int_0^t \langle i_C(\xi) i_C(\zeta) \rangle\, d\xi\, d\zeta$$

and Rice[3] has shown that

$$\langle i_C(\xi) i_C(\zeta) \rangle = \langle i_S(\xi) i_S(\zeta) \rangle = \int_0^\infty w(f) \cos 2\pi[(f - f_0)(\zeta - \xi)]\, df \quad (5)$$

where

$$w(f) = 4 \int_0^\infty \psi(\tau) \cos 2\pi f\tau\, d\tau$$

is the power spectrum of the $i(t)$ process. Thus

$$\langle I_C^2(t) \rangle = \langle I_S^2(t) \rangle = \int_0^\infty w(f) \left[\frac{\sin \pi t(f - f_0)}{\pi(f - f_0)}\right]^2 df. \quad (6)$$

Also,

$$\langle i_S(\xi) i_C(\zeta) \rangle = \int_0^\infty w(f) \sin 2\pi[(f - f_0)(\zeta - \xi)]\, df$$

from which it follows that

$$\langle I_C(t) I_S(t) \rangle = \int_0^t \int_0^t \langle i_S(\xi) i_C(\zeta) \rangle\, d\xi\, d\zeta = 0$$

and hence

$$\langle I^2(t) \rangle = \langle I_C^2(t) \rangle \cos^2 \omega_0 t + \langle I_S^2(t) \rangle \sin^2 \omega_0 t$$

is equal to the common value of $\langle I_C^2(t) \rangle$ and $\langle I_S^2(t) \rangle$ as given by (6). For large values of t, $\langle I^2(t) \rangle$ is asymptotic to $t \cdot w(f_0)$.

It is also of interest to compute the correlation function of the outputs of two ICI of different tuned frequencies. Suppose, then that $i(t)$ is a stationary Gaussian process of mean zero and correlation function $\psi(\tau)$. The input to an ICI of resonant frequency ω_r is

$$i(t) = i_C(t) \cos \omega_r t + i_S(t) \sin \omega_r t,$$

and the output is

$$I_r(t) = \left[\int_0^t i_C(\xi)\, d\xi\right] \cos \omega_r t + \left[\int_0^t i_S(\xi)\, d\xi\right] \sin \omega_r t.$$

Similarly, the input to an ICI of frequency ω_s may be written

$$i(t) = i_C'(t) \cos \omega_s t + i_S'(t) \sin \omega_s t$$

with output

$$I_s(t) = \left[\int_0^t i_C'(\xi)\, d\xi\right] \cos \omega_s t + \left[\int_0^t i_S'(\xi)\, d\xi\right] \sin \omega_s t.$$

Following Rice it is easy to see that

$$\langle i_C(\xi) i_C'(\zeta) \rangle = \langle i_S(\xi) i_S'(\zeta) \rangle = \int_0^\infty w(f) \cos 2\pi[(f - f_r)\xi - (f - f_s)\zeta]\, df$$

and

$$-\langle i_S(\xi) i_C'(\zeta) \rangle = \langle i_C(\xi) i_S'(\zeta) \rangle = \int_0^\infty w(f) \sin 2\pi[(f - f_r)\xi - (f - f_s)\zeta]\, df$$

where $w(f)$ is the (stationary) power spectrum of $i(t)$.

[2] S. O. Rice, "Mathematical analysis of random noise," *Bell Sys. Tech. J.*, vol. 24, pp. 46-156; January, 1945. See p. 75.
[3] *Ibid.*, p. 77.

We define the cross correlation of $I_r(t)$ and $I_s(t)$ (at the same instant) as

$$\Psi_{rs}(t) = \langle I_r(t) I_s(t) \rangle.$$

Then it follows as before that

$$\Psi_{rs}(t) = \cos \pi t(f_s - f_r) \cdot \int_0^\infty w(f) \left[\frac{\sin \pi t(f - f_r)}{\pi(f - f_r)} \right] \left[\frac{\sin \pi t(f - f_s)}{\pi(f - f_s)} \right] df$$

which vanishes for values of t which make $\pi t(f_s - f_r)$ an odd multiple of $\pi/2$ with no restrictions on the form of the spectral density of $i(t)$.

C. Cumulative Distribution of the Envelope of ICI

If we have a bank of n ICI tuned to the frequencies $\omega_1, \cdots, \omega_n$, we wish to compute the cumulative distribution of the *envelopes* of the outputs. Thus, representing the output of the rth ICI as

$$I_r(t) = I_{Cr}(t) \cos \omega_r t + I_{Sr}(t) \sin \omega_r t, \qquad r = 1, 2, \cdots, n \tag{7}$$

the envelope is defined as

$$R_r(t) = \sqrt{I_{Cr}^2(t) + I_{Sr}^2(t)}. \tag{8}$$

If the joint fr.f. of the R's is denoted by $p(R_1, \cdots, R_n)$, then the cumulative distribution C is

$$C = \int_0^\tau \cdots \int_0^\tau p(R_1, \cdots, R_n) \, dR_1 \cdots dR_n. \tag{9}$$

Suppose then that the input to each ici is statistically stationary Gaussian noise $i(t)$ with mean zero, correlation function $\psi(\tau)$, and power spectrum $w(f)$. Thus $I_{C1}, I_{S1}, I_{C2}, I_{S2}, \cdots, I_{Cn}, I_{Sn}$ have a joint $2n$-dimensional Gaussian distribution. As in Part B, it is seen that

$$\Psi_{rs}(t) = \langle I_{Cr} I_{Cs} \rangle = \langle I_{Sr} I_{Ss} \rangle$$

$$= \int_0^t \int_0^t d\xi \, d\zeta \int_0^\infty w(f) \cos 2\pi[(f - f_r)\xi - (f - f_s)\zeta] \, df$$

$$= \cos \pi t(f_s - f_r) \int_0^\infty w(f) F(f - f_r) F(f - f_s) \, df \tag{10}$$

and

$$\Phi_{rs}(t) = \langle I_{Cr} I_{Ss} \rangle = -\langle I_{Sr} I_{Cs} \rangle$$

$$= \int_0^t \int_0^t d\xi \, d\zeta \int_0^\infty w(f) \sin 2\pi[(f - f_r)\xi - (f - f_s)\zeta] \, df$$

$$= \sin \pi t(f_s - f_r) \int_0^\infty w(f) F(f - f_r) F(f - f_s) \, df \tag{11}$$

where

$$F(x) = \frac{\sin \pi t x}{\pi x}. \tag{12}$$

The correlation matrix of the $2n$-dimensional Gaussian distribution is thus

$$M = \left\| \begin{matrix} \Psi_{11} & 0 & \Psi_{12} & \Phi_{12} & \cdots & \Psi_{1n} & \Phi_{1n} \\ 0 & \Psi_{11} & -\Phi_{12} & \Psi_{12} & \cdots & -\Phi_{1n} & \Psi_{1n} \\ \Psi_{21} & \Phi_{21} & \Psi_{22} & 0 & \cdots & \Psi_{2n} & \Phi_{2n} \\ -\Phi_{21} & \Psi_{21} & 0 & \Psi_{22} & \cdots & -\Phi_{2n} & \Psi_{2n} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \Psi_{n1} & \Phi_{n1} & \Psi_{n2} & \Phi_{n2} & \cdots & \Psi_{nn} & 0 \\ -\Phi_{n1} & \Psi_{n1} & -\Phi_{n2} & \Psi_{n2} & \cdots & 0 & \Psi_{nn} \end{matrix} \right\| \tag{13}$$

with respect to the variables $I_{C1}, I_{S1}, \cdots, I_{Cn}, I_{Sn}$ in that order. Note that $\Psi_{rs} = \Psi_{sr}$ and $\Phi_{rs} = -\Phi_{sr}$. Calling $q(I_{C1}, \cdots, I_{Sn})$ the joint $2n$-dimensional distribution, and letting

$$I_{Cr} = R_r \cos \theta_r, \qquad I_{Sr} = R_r \sin \theta_r, \qquad r = 1, 2, \cdots, n$$

the joint fr.f. of the R's and θ's is

$$f(R_1, \theta_1, \cdots, R_n, \theta_n) = J\left(\frac{I_{C1}, I_{S1}, \cdots, I_{Cn}, I_{Sn}}{R_1, \theta_1, \cdots, R_n, \theta_n} \right) \cdot q(I_{C1}, I_{S1}, \cdots, I_{Cn}, I_{Sn}) \tag{14}$$

where J is the Jacobian of the transformation from the I_C, I_S variables to the R and θ variables. The fr.f. of the R's is thus given by the marginal distribution

$$p(R_1, \cdots, R_n) = \int_0^{2\pi} \cdots \int_0^{2\pi} f(R_1, \theta_1, \cdots, R_n, \theta_n) \, d\theta_1 \cdots d\theta_n \tag{15}$$

which allows us to calculate the cumulative distribution C of (9).

In order to compute $p(R_1, \cdots, R_n)$, we resort to the following technique. The fr.f $p(R_1, \cdots, R_n)$ depends on the Ψ_{rs} and the Φ_{ij} as can be seen from (13). Therefore, consider $p(R_1, \cdots, R_n)$ as depending on the parameters G_{rs} where

$$\Phi_{rs}(t) = \sin \pi t(f_s - f_r) G_{rs}, \qquad \Psi_{rs}(t) = \cos \pi t(f_s - f_r) G_{rs};$$

that is,

$$G_{rs} = \int_0^\infty w(f) F(f - f_r) F(f - f_s) \, df, \qquad r, s = 1, 2, \cdots, n. \tag{16}$$

Because $G_{rs} = G_{sr}$ only the cases $s > r$ need be considered. Thus, we may write

$$p(R_1, \cdots, R_n) = g(\mathbf{R}, \mathbf{G})$$

where $\mathbf{R}$ is the n tuple $R_1, \cdots, R_n$ and $\mathbf{G}$ is the $\frac{1}{2}n(n-1)$ dimensional parameter G_{rs}, $s > r$; $s, r = 1, \cdots, n$. Now expand $g(\mathbf{R}, \mathbf{G})$ in a multiple power series,

$$g(\mathbf{R}, \mathbf{G}) = g(\mathbf{R}, \mathbf{0}) + \sum_{\substack{r,s=1 \\ s>r}}^{n} \frac{\partial}{\partial G_{rs}} g(\mathbf{R}, \mathbf{0}) G_{rs} + \frac{1}{2!} \sum_{\substack{r,s=1 \\ s>r}}^{n} \sum_{\substack{i,j=1 \\ j>i}}^{n} \frac{\partial^2 g(\mathbf{R}, \mathbf{0})}{\partial G_{rs} \, \partial G_{ij}} G_{rs} G_{ij} + \cdots. \tag{17}$$

This is a valid representation for $g(\mathbf{R}, \mathbf{G})$ within the radius of convergence of the power series.

The value of the above representation is the following. If $\mathbf{G} = \mathbf{0}$, then the correlation matrix M of (13) is purely diagonal. Hence $g(\mathbf{R}, \mathbf{0})$ is simply the product of the one-dimensional marginal distributions:

$$g(\mathbf{R}, \mathbf{0}) = p_1(R_1)p_2(R_2) \cdots p_n(R_n).$$

The marginal distributions $p_i(R_i)$ are Rayleigh:

$$p_i(R_i) = \frac{R_i}{\Psi_{ii}} e^{-R_i^2/2\Psi_{ii}}. \tag{18}$$

Now to compute $\partial g\,(\mathbf{R}, \mathbf{0})/\partial G_{rs}$, first set every G_{ij} except G_{rs} equal to zero, differentiate with respect to G_{rs}, and then set *it* equal to zero. If we let $g(\mathbf{R}, \hat{G}_{rs})$ represent $g(\mathbf{R}, \mathbf{G})$ with every G_{ij} except G_{rs} set equal to zero then

$$g(\mathbf{R}, \hat{G}_{rs}) = p(R_r, R_s)\Pi_{rs} \tag{19}$$

where Π_{rs} is the product of the *a priori* distributions $p_1(R_1), \cdots, p_n(R_n)$ except $p_r(R_r)$ and $p_s(R_s)$. Thus, the problem is reduced to that of calculating $p(R_r, R_s)$, and

$$\frac{\partial}{\partial G_{rs}} g(\mathbf{R}, \mathbf{0}) = \left(\frac{\partial}{\partial G_{rs}} p(R_r, R_s)\right)_{G_{rs}=0} \Pi_{rs}. \tag{20}$$

The computation of $\partial^2 g(\mathbf{R}, \mathbf{0})/\partial G_{rs}\partial G_{ij}$ is somewhat more difficult. If the pair (r, s) is the same as (i, j), the formulation of (19) may be used to compute $\partial^2 g(\mathbf{R},\mathbf{0})/\partial G_{rs}^2$. If (r, s) is distinct from (i, j), then

$$\frac{\partial^2 g(\mathbf{R}, \mathbf{0})}{\partial G_{rs}\,\partial G_{ij}} = \left(\frac{\partial}{\partial G_{rs}} p(R_r, R_s)\right)_{\mathbf{G}=\mathbf{0}} \cdot \left(\frac{\partial}{\partial G_{ij}} p(R_i, R_j)\right)_{\mathbf{G}=\mathbf{0}} \Pi_{rsij} \tag{21}$$

where Π_{rsij} is defined in the expected fashion. Finally, we must consider the cases where $r = i$ and $s \neq j$ or $r \neq i$ and $s = j$. In these cases we must compute $p(R_r, R_i, R_j)$ (assuming $s = j$). However, referring to the correlation matrix M of (13) it can be shown by a lengthy but elementary calculation that

$$\left(\frac{\partial^2}{\partial G_{rj}\,\partial G_{ij}} p(R_r, R_i, R_j)\right)_{\mathbf{G}=\mathbf{0}} = 0, \tag{22}$$

[cf. (15)].

Now turn to the computation of $p(R_r, R_s)$. For this case the correlation matrix becomes[4]

$$M = \begin{Vmatrix} G_{rr} & 0 & G_{rs}\cos x & G_{rs}\sin x \\ 0 & G_{rr} & -G_{rs}\sin x & G_{rs}\cos x \\ G_{rs}\cos x & -G_{rs}\sin x & G_{ss} & 0 \\ G_{rs}\sin x & G_{rs}\cos x & 0 & G_{ss} \end{Vmatrix} \tag{23}$$

where

$$x = \pi t(f_s - f_r). \tag{24}$$

If A_{rs} is defined as

$$A_{rs} = \begin{vmatrix} G_{rr} & G_{rs} \\ G_{rs} & G_{ss} \end{vmatrix} \tag{25}$$

then an elementary calculation shows that the determinant of M is $\det M = A_{rs}^2$ and the cofactors of this symmetric matrix are

$$\begin{aligned} &M_{11} = M_{22} = A_{rs}G_{ss} && M_{33} = M_{44} = A_{rs}G_{rr} \\ &M_{12} = M_{34} = 0 && M_{23} = -M_{14} = A_{rs}G_{rs}\sin x \\ & && M_{24} = M_{13} = -A_{rs}G_{rs}\cos x. \end{aligned}$$

The joint distribution $q(I_{Cr}, I_{Sr}, I_{Cs}, I_{Ss})$ is thus

$$\begin{aligned} q(I_{Cr}, I_{Sr}, I_{Cs}, I_{Ss}) &= \frac{1}{4\pi^2 A_{rs}} \exp\left\{-\frac{1}{2A_{rs}}\left[G_{ss}(I_{Cr}^2 + I_{Sr}^2) + G_{rr}(I_{Cs}^2 + I_{Ss}^2)\right.\right. \\ &\quad - 2G_{rs}\sin x(I_{Cr}I_{Ss} - I_{Sr}I_{Cs}) \\ &\quad \left.\left. - 2G_{rs}\cos x(I_{Sr}I_{Ss} + I_{Cr}I_{Cs})\right]\right\}. \end{aligned}$$

The Jacobian appearing in (14) is $J = R_rR_s$ and hence the marginal distribution $p(R_r, R_s)$ is

$$\begin{aligned} p(R_r, R_s) &= \frac{R_rR_s}{4\pi^2 A_{rs}} \exp\left[-\frac{1}{2A_{rs}}(G_{ss}R_r^2 + G_{rr}R_s^2)\right] \\ &\quad \times \int_0^{2\pi}\int_0^{2\pi} \exp\left\{\frac{G_{rs}R_rR_s}{A_{rs}}[\sin x \sin(\theta_s - \theta_r) \right. \\ &\quad \left. + \cos x \cos(\theta_s - \theta_r)]\right\} d\theta_r\, d\theta_s \\ &= \frac{R_rR_s}{A_{rs}} \exp\left[-\frac{1}{2A_{rs}}(G_{ss}R_r^2 + G_{rr}R_s^2)\right] I_0\left(\frac{G_{rs}R_rR_s}{A_{rs}}\right) \end{aligned} \tag{26}$$

where I_0 is the Bessel function of the first kind and order zero with purely imaginary argument.

One now sees immediately that

$$\left(\frac{\partial}{\partial G_{rs}} p(R_r, R_s)\right)_{G_{rs}=0} = 0 \tag{27}$$

and

$$\left(\frac{\partial^2}{\partial G_{rs}^2} p(R_r, R_s)\right)_{G_{rs}=0} = \frac{2p_r(R_r)p_s(R_s)}{\Psi_{rr}\Psi_{ss}}\left(1 - \frac{R_r^2}{2\Psi_{rr}}\right)\left(1 - \frac{R_s^2}{2\Psi_{ss}}\right) \tag{28}$$

where we recall that $G_{rr} = \Psi_{rr}$. Eq. (17) then becomes

[4] For the case $G_{rr} = G_{ss}$ this is similar to a result of D. Middleton, "Some general results in the theory of noise through non-linear devices," *Quart. Appl. Math.*, vol. 5, pp. 445-498; January, 1948. In particular, see pp. 469-472.

$$p(R_1, \cdots, R_n) = \prod_{i=1}^{n} p_i(R_i) + \frac{1}{2!} \sum_{\substack{r,s=1\\ s>r}}^{n} \frac{\partial^2}{\partial G_{rs}^2} g(\mathbf{R}, \mathbf{0}) G_{rs}^2 + \cdots$$

$$= \prod_{i=1}^{n} p_i(R_i) + \prod_{i=1}^{n} p_i(R_i) \cdot \sum_{\substack{r,s=1\\ s>r}}^{n} \frac{G_{rs}^2}{\Psi_{rr}\Psi_{ss}} \left(1 - \frac{R_r^2}{2\Psi_{rr}}\right)\left(1 - \frac{R_s^2}{2\Psi_{ss}}\right) + \cdots. \quad (29)$$

The cumulative distribution $C(\tau)$ of (9) is thus

$$C(\tau) = \prod_{i=1}^{n} H_i(\tau) + \frac{1}{4} \tau^4 \sum_{\substack{r,s=1\\ s>r}}^{n} \frac{G_{rs}^2}{\Psi_{rr}^2 \Psi_{ss}^2} \cdot \exp\left[-\frac{\tau^2}{2}\left(\frac{1}{\Psi_{rr}} + \frac{1}{\Psi_{ss}}\right)\right] \prod_{\substack{i=1\\ i\neq r\\ i\neq s}}^{n} H_i(\tau) + \cdots \quad (30)$$

where

$$H_r(\tau) = \int_0^\tau p_r(R_r)\, dR_r = \int_0^\tau \frac{R_r}{\Psi_{rr}} e^{-R_r^2/2\Psi_{rr}}\, dR_r = (1 - e^{-\tau^2/2\Psi_{rr}}). \quad (31)$$

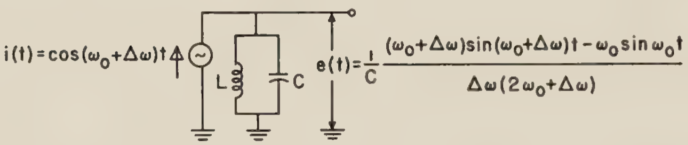


Fig. 3—Lossless tank circuit used as an approximate ideal coherent integrator.

D. An Approximate Ideal Coherent Integrator

An example of a physical device which performs almost as an ICI is furnished by a lossless tank circuit as illustrated in Fig. 3. The ratio of the output voltage to the input current is

$$Z(s) = \frac{sL}{1 + s^2 LC}$$

where $s = j\omega$. We call the inverse Laplace transform of $Z(s)$ the impulsive response of the network. It is

$$h(t) = \frac{1}{C} \cos \omega_0 t, \qquad t > 0$$

$$h(t) = 0, \qquad t < 0$$

where $\omega_0^2 = 1/LC$. If we let[5] $C = 1/2$, then the impulsive response becomes

$$h(t) = 2 \cos \omega_0 t, \qquad t > 0$$
$$= 0, \qquad t < 0.$$

[5] The output of the network for an input $i(t) = \sin \omega_0 t$ is $e(t) = (t/2C) \sin \omega_0 t$. Since we want to approximate an ICI as closely as possible, we choose C so that the coefficient of t is one, that is, $1/2C = 1$ or $C = \frac{1}{2}$.

By the superposition theorem, the voltage $e(t)$ across the tank circuit due to a current $i(t)$ applied at $t = 0$ is

$$e(t) = \int_0^t h(\xi) i(t - \xi)\, d\xi.$$

For example, if $i(t) = \sin \omega_0 t$, then $e(t) = t \sin \omega_0 t$. If $i(t) = t \sin \omega_0 t$, then

$$e(t) = \frac{1}{2} t^2 \sin \omega_0 t + \frac{\sin \omega_0 t}{2\omega_0^2} - \frac{t \cos \omega_0 t}{2\omega_0} \doteq \frac{1}{2} t^2 \sin \omega_0 t$$

for large ω_0 and/or t. If $i(t) = \cos(\omega_0 + \Delta\omega)t$, then

$$e(t) = \frac{2(\omega_0 + \Delta\omega) \sin (\omega_0 + \Delta\omega)t - 2\omega_0 \sin \omega_0 t}{\Delta\omega(2\omega_0 + \Delta\omega)}.$$

When the deviation $\Delta\omega$ of the applied frequency from the resonant frequency ω_0 is small compared with the resonant frequency; that is, $|\Delta\omega| \ll \omega_0$, then to a first-order approximation we may write

$$e(t) = \frac{2\omega_0 \sin(\omega_0 + \Delta\omega)t - 2\omega_0 \sin \omega_0 t}{2\omega_0 \Delta\omega}$$

$$= \frac{\sin \frac{\Delta\omega t}{2}}{\frac{\Delta\omega}{2}} \cos\left(\omega_0 t + \frac{\Delta\omega t}{2}\right)$$

which is identical with (3).

E. Statistical Properties of an Approximate ICI

Suppose that stationary Gaussian noise $i(t)$ with mean zero and correlation function $\psi(\tau)$ is applied to an approximate ICI with impulse response

$$h(t) = 2 \cos \omega_0 t, \qquad t > 0$$
$$= 0, \qquad t < 0.$$

Then the output is, by the superposition theorem,

$$I(t) = \int_0^t h(t - \xi) i(\xi)\, d\xi.$$

It is readily seen that $\langle I(t) \rangle = 0$. The variance of $I(t)$ may be computed from the formula

$$\langle I^2(t) \rangle = \int_0^t \int_0^t h(t - \xi) h(t - \zeta) \langle i(\xi) i(\zeta) \rangle\, d\xi\, d\zeta.$$

Following the notation of Lampard,[6]

$$\langle I^2(t) \rangle = 4 \int_0^t \phi(t - x; x) \psi(x)\, dx \quad (32)$$

where $\phi(t; x) = \frac{1}{2} \int_0^t h(\xi) h(\xi + x)\, d\xi$ is the time-dependent filter correlation function. We readily compute

$$\phi(t - x; x) = (t - x) \cos \omega_0 x + \frac{1}{\omega_0} \sin \omega_0 (t - x) \cos \omega_0 t. \quad (33)$$

[6] D. G. Lampard, "The response of linear networks to suddenly applied stationary random noise," IRE Trans., vol. CT-2, pp. 49-57; March, 1955.

If $w(f)$ is the power spectrum of the $i(t)$ process, then

$$\psi(x) = \int_0^\infty w(f) \cos \omega x \, df$$

and

$$\langle I^2(t)\rangle = \int_0^\infty w(f)[F^2(f + f_0) + F^2(f - f_0)]\, df + 2 \cos 2\pi f_0 t \int_0^\infty w(f)F(f + f_0)F(f - f_0)\, df \tag{34}$$

where $F(x) = (\sin \pi tx)/\pi x$, [cf. (12)].

If $w(f)$ is white, $w(f) = w_0$ and

$$\langle I^2(t)\rangle = w_0\left[t + \frac{\sin 2\omega_0 t}{2\omega_0}\right]. \tag{35}$$

This formula could also have been more readily obtain by noting that for $w(f) = w_0$, $\psi(x) = \frac{1}{2}w_0\, \delta(x)$ and hence (32) yields

$$\langle I^2(t)\rangle = w_0\phi(t; 0). \tag{36}$$

Applying (33) then yields (35).

If $i(t)$ is now introduced into two distinct approximate ICI, say of resonant frequencies ω_r and ω_s, then their outputs will be

$$I_r(t) = \int_0^t h_r(\xi)i(t - \xi)\, d\xi$$

and

$$I_s(t) = \int_0^t h_s(\xi)i(t - \xi)\, d\xi$$

respectively, where

$$\begin{aligned} h_r(t) &= 2 \cos \omega_r t, \\ h_s(t) &= 2 \cos \omega_s t. \end{aligned} \tag{37}$$

Again following Lampard, define $\Psi_{rs}(t; \tau) = \langle I_r(t)I_s(t + \tau)\rangle$ as the time-dependent correlation function. Since we are interested only in the outputs at the same time t, we write $\Psi_{rs}(t) = \langle I_r(t)I_s(t)\rangle$. One then finds by a direct calculation that

$$\begin{aligned} \Psi_{rs}(t) = {} & \cos \pi t(f_r + f_s) \int_0^\infty w(f)[F(f + f_s)F(f - f_r) + F(f + f_r)F(f - f_s)]\, df \\ & + \cos \pi t(f_r - f_s) \int_0^\infty w(f)[F(f + f_s)F(f + f_r) + F(f - f_r)F(f - f_s)]\, df \end{aligned} \tag{38}$$

where $F(x)$ has been defined in (12). Note that if ω_r and ω_s are commensurate frequencies, $\Psi_{rs}(t)$ vanishes for suitable values of t—again with no restrictions on the form of the power spectrum $w(f)$. In particular, if $w(f) = w_0$, then

$$\Psi_{rs}(t) = w_0\left[\frac{\sin (\omega_r + \omega_s)t}{\omega_r + \omega_s} + \frac{\sin (\omega_s - \omega_r)t}{\omega_s - \omega_r}\right],$$

which could also have been easily obtained by setting $\psi(x) = \frac{1}{2}w_0\, \delta(x)$.

III. Improvement of the Signal-To-Noise Ratio

The signal to be detected is assumed to be a constant amplitude, constant frequency sinusoid $x(t)$ which is additively combined with stationary Gaussian noise, $i(t)$. We shall assume that $i(t)$ has mean zero, correlation function $\psi(\tau)$ and power spectrum $\omega(f)$.

The input to an ICI of resonant frequency ω_0 is $x(t) + i(t)$. As in Section II-A, the constant amplitude signal $x(t)$ may be written as

$$x(t) = a \cos \omega_0 t + b \sin \omega_0 t \tag{39}$$

where a and b describe the envelope, and following Rice (cf. Section II-B) $i(t)$ may be written as

$$i(t) = i_C(t) \cos \omega_0 t + i_S(t) \sin \omega_0 t \tag{40}$$

where

$$\langle i\rangle = \langle i_C\rangle = \langle i_S\rangle = 0$$

$$\langle i^2\rangle = \langle i_C^2\rangle = \langle i_S^2\rangle = \psi_0$$

and i_C and i_S are independent Gaussian processes. The input to the ICI of tuned frequency ω_0 is thus

$$e_1(t) = x(t) + i(t) = [a + i_C(t)] \cos \omega_0 t + [b + i_S(t)] \sin \omega_0 t \tag{41}$$

and by definition the output is

$$e_2(t) = \cos \omega_0 t \int_0^t [a + i_C(\xi)]\, d\xi + \sin \omega_0 t \int_0^t [b + i_S(\xi)]\, d\xi. \tag{42}$$

Define the input signal power P_{S1} as

$$P_{S1} = \tfrac{1}{2}(a^2 + b^2) \tag{43}$$

and the output signal power P_{S2} as

$$P_{S2} = \frac{1}{2}\left[\left(\int_0^t a\, d\xi\right)^2 + \left(\int_0^t b\, d\xi\right)^2\right] = \frac{1}{2}\, t^2(a^2 + b^2). \tag{44}$$

The input noise power P_{N1} is defined as

$$P_{N1} = \tfrac{1}{2}[\langle i_C^2\rangle + \langle i_S^2\rangle] = \psi_0 \tag{45}$$

and the output noise power P_{N2} is defined as

$$P_{N2} = \frac{1}{2}\left\{\left\langle\left[\int_0^t i_C(\xi)\, d\xi\right]^2\right\rangle + \left\langle\left[\int_0^t i_S(\xi)\, d\xi\right]^2\right\rangle\right\} = \int_0^\infty w(f)F^2(f - f_0)\, df \tag{46}$$

by (6) where $F(x) = (\sin \pi tx)/\pi x$ [cf. (12)].

The input signal-to-noise ratio ρ_1 is therefore

$$\rho_1 = \frac{P_{S1}}{P_{N1}} = \frac{a^2 + b^2}{2\psi_0} \tag{47}$$

and the output signal-to-noise ratio ρ_2 is

$$\rho_2 = \frac{P_{S2}}{P_{N2}} = \frac{t^2(a^2 + b^2)}{2 \int_0^\infty w(f)F^2(f - f_0)\, df}. \qquad (48)$$

Clearly, if $\rho_2 > \rho_1$ our process has resulted in a signal enhancement. Thus a measure of improvement is given by the ratio of the signal-to-noise ratios; that is, by

$$\lambda = \frac{\rho_2}{\rho_1} = \frac{t^2\psi_0}{\int_0^\infty w(f)F^2(f - f_0)\, df}. \qquad (49)$$

If we make the change of variable $\pi t(f - f_0) = x$, then λ becomes

$$\lambda = \frac{\pi t\psi_0}{\int_{-\pi t f_0}^\infty w\left(\frac{x}{\pi t} + f_0\right)\left(\frac{\sin x}{x}\right)^2 dx} \qquad (50)$$

which, for large values of t, is asymptotic to

$$\lambda \sim \frac{t\psi_0}{w(f_0)}. \qquad (51)$$

Thus λ increases linearly with time, and the longer the integration period t, the greater the signal-to-noise ratio improvement.

As a practical example, consider the case where the noise has a band-limited power spectrum. That is,

$$w(f) = w_0, \qquad 0 \leqq f < B$$

$$w(f) = 0, \qquad B < f.$$

Then if the integration time is T, the signal-to-noise ratio improvement factor λ is kBT where

$$k^{-1} = \frac{1}{\pi}\int_{-\pi T f_0}^{\pi T(B - f_0)} \left(\frac{\sin x}{x}\right)^2 dx.$$

For almost all values of the coherent integrators' tuned frequencies f_0 except those near the edges of the frequency band B, the value of k is essentially one.

IV. The Multichannel Problem

The signal-to-noise improvement factor λ is directly proportional to the duration of the integration period [cf. (51)] while the effective bandwidth of an ICI decreases as the integration period increases [cf. (4)]. In certain important radar applications, it is known *a priori* that if the signal occurs it must fall within a known frequency range which shall be designated the "signal band," B. We treat the case where the signal band and noise band are identical. For example, in the case where the only cause of frequency difference between the transmitted and received signals is the Doppler effect produced by the target's velocity, the signal band can be specified from a knowledge of the range of possible target velocities. However, since the exact echo frequency is not known *a priori*, we must be prepared to process an echo which occurs anywhere within the signal band.

The number of integration channels, n, required for the entire signal band increases as the integration period is lengthened. Each integration channel can independently cause a false alarm, although only the particular channel in which the signal appears can cause a true alarm. Therefore, the question arises as to whether it is desirable to lengthen the coherent integration period T to improve the signal-to-noise ratio, when so doing requires an increase in the number of integration channels n and a consequent increase in the opportunity for a false alarm.

Of course, the problem is to maximize the detection probability $1 - \beta$, or equivalently, to minimize the missed-signal probability β. Suppose that the signal occurs in the rth channel. Then the output of this channel is of the form[7]

$$[I_C(t) + at] \cos \omega_r t + [I_S(t) + bt] \sin \omega_r t. \qquad (52)$$

Its envelope R is then the square root of the sum of the squares of two independently normally distributed random variables with common variance $\Psi_{rr}(t)$ and means at and bt respectively. The fr.f. of R is then

$$p(R) = \frac{R}{\Psi_{rr}} e^{-[R^2+t^2(a^2+b^2)]/2\Psi_{rr}} I_0\left(\frac{Rt\sqrt{a^2 + b^2}}{\Psi_{rr}}\right) \qquad (53)$$

and the detection probability is

$$1 - \beta = \int_\tau^\infty p(R)\, dR \qquad (54)$$

where τ is the acceptance threshold.

The false alarm probability α is related to the threshold τ by

$$\alpha = 1 - \int_0^\tau \cdots \int_0^\tau p(R_1, \cdots, R_n)\, dR_1 \cdots dR_n \qquad (55)$$

where $p(R_1, \cdots, R_n)$ is the joint n-dimensional distribution of the envelopes of the outputs of the coherent integrators in the absence of a signal. Theoretically, if we choose α and n, then τ is determined by (55) as a function of the integration time T, and thus β is determined by (54). The problem, formulated mathematically, is to determine the integration time T, threshold τ, and number of channels n such that β is minimized.

These quantities cannot be specified independently. The value of n is determined by how far down on the skirt of the integrator's response characteristic, given by (4), the signal frequency will be permitted to fall before coming into the coverage of the adjacent integration

[7] Eq. (52) is written on the assumption that the input signal is of the form $a \cos \omega_r t + b \sin \omega_r t$; that is, the frequency of the signal is exactly the tuned frequency of the r th ICI. If the signal frequency is actually $\omega_r + \delta\omega$, then the terms $t^2(a^2 + b^2)$ in (53) must be replaced by

$$\left[\frac{\sin \frac{\delta\omega t}{2}}{\frac{\delta\omega}{2}}\right]^2 (a^2 + b^2).$$

However, for small $\delta\omega$, $(a^2 + b^2)t^2$ is a good approximation to the above expression.

channel. For the present analysis, we shall assume that the integration channels overlap at a frequency such that $\Delta\omega T = \pi/2$. This causes the amplitude of $E(T, \Delta\omega)$ to be equal to minus 0.912 db at the cross over frequency. The largest channel separation that reasonably might be chosen is one that causes $\Delta\omega T = 2\pi$, since adjacent channels would then overlap where $E(T, \Delta\omega) = 0$. However, this choice is not very satisfactory because a substantial portion of the coverage of an integration channel would include frequencies for which $E(T, \Delta\omega)$ is far from unity. The choice of $\Delta\omega T = \pi/2$ is probably the most satisfactory since the integrator's response is almost uniform in the frequency band it is assigned to cover. For this particular choice, it is found the number of integration channels required is

$$n = 2BT \tag{56}$$

where B is the signal band in cps and T the integration time in seconds. Thus $T = 1/2f_1$ where $f_1 = 2\Delta f$ is the separation in cps between adjacent channels.

If it is assumed that the input noise has a band-limited white spectrum (of width B cps) then $\Psi_{rr}(t)$ can be determined. For in this case,

$$\begin{aligned} w(f) &= w_0, \qquad 0 \leqq f < B \\ w(f) &= 0, \qquad B < f \end{aligned} \tag{57}$$

where $w(f)$ is the power spectrum of the input noise. It is then not difficult to see, that except for very small values of t and ω_r, that with but negligible error

$$\Psi_{rr}(T) = w_0 T, \tag{58}$$

where T is the coherent integration time (cf. Section III).

If $T = 1/2f_1$, then the G_{rs} terms of (30) will be small if $|r - s|$ is large. If $|r - s|$ is small and odd, G_{rs} will not be negligible compared with Ψ_{rr}. However, numerical calculations seem to verify the conclusion that even for these cases and large n the second term of the series is small compared with the first. Hence to a good approximation [cf. (55)],

$$\alpha = 1 - (1 - e^{-\tau^2/2w_0T})^n$$

or

$$\tau = \sqrt{-2w_0T \log [1 - (1 - \alpha)^{1/n}]}. \tag{59}$$

Substituting this value of τ in (54) and recalling that $n = 2BT$ by (56) we have, after a change of variable that[8]

$$\beta = \int_0^\xi ue^{-(u^2+2Qn)/2} I_0(u\sqrt{2Qn})\, du \tag{60}$$

where

$$Q = \frac{T(a^2 + b^2)}{2nw_0} = \frac{1}{n}\rho_2 = \frac{1}{2}\rho_1$$

[cf. (47, 48)], and

$$\xi = \sqrt{-2 \log [1 - (1 - \alpha)^{1/n}]}. \tag{61}$$

Thus, the problem is reduced to the determination of one parameter, namely n, such that β is minimized. In the Appendix, β is examined as a function of n. It turns out that β is *not* a monotonic function of n. In fact,

$$\lim_{n\to 0} \beta = 0 = \lim_{n\to\infty} \beta \tag{62}$$

and β has a unique maximum at the only positive root of

$$\frac{d\xi}{db} = \frac{I_1(b\xi)}{I_0(b\xi)} \tag{63}$$

where $b = 2Qn$. For example, if the over-all false alarm probability α is 0.1 and the initial signal-to-noise ratio $\rho_1 = 0.01 = -20$ db, then the worst number of channels to use is 53 or 54.

However, for the radar applications in which we are interested, (cf. Figs. 4–11, pp. 246–247) maximum value of β occurs at a value of n less than one. Thus, since the number of channels must be an integer, the missed-signal probability β is a monotonic function of n. The best results are therefore obtainable by using as large a number of coherent integrators as is practical. Numerical results are presented graphically in Figs. 4 through 11. The over-all false alarm probability α is constant for each graph, while the initial signal-to-noise ratio ρ_1 varies from curve to curve, with the missed-signal probability β plotted vs the number of channels n.

Appendix

The fundamental formula for the missed-signal probability is given by (60) as

$$\beta = \int_0^\xi ue^{-(u^2+b^2)/2} I_0(bu)\, du \tag{64}$$

where

$$b = \sqrt{2Qn}, \qquad \xi = \sqrt{-2 \log (1 - a^{1/n})} \tag{65}$$

and $a = 1 - \alpha$.

The integrand of (64) is non-negative and hence, Case 1,

$$\beta \geqq 0.$$

Since

$$\begin{aligned} p(u) &= ue^{-(u^2+b^2)/2} I_0(bu), \qquad u > 0 \\ &= 0, \qquad u < 0 \end{aligned}$$

is a frequency function it is seen that for b fixed and finite, Case 2,

$$\int_0^\infty p(u)\, du = 1.$$

The first nontrivial properties of β that will be demonstrated are, Case 3,

$$\lim_{n\to 0} \beta = 0$$

[8] This is essentially the Q function of J. I. Marcum, "A Statistical Theory of Target Detection by Pulsed Radar: Mathematical Appendix," The RAND Corp., Santa Monica, Calif., RM-753; July 1, 1948, reissued April 25, 1952. In Marcum's notation, our β is $1 - Q(\sqrt{2Qn}, \xi)$.

and, Case 4,

$$\lim_{n\to\infty} \beta = 0.$$

From (65), $\lim_{n\to 0} b = 0$ and $\lim_{n\to 0} \xi = 0$. Thus $\beta(0) = 0$ since the upper limit of the integral, ξ, becomes zero while the integrand remains bounded, (in fact it is precisely $ue^{-u^2/2}$). This proves Case 3.

A sequence of inequalities will be deduced in order to establish Case 4. From Stirling's formula for the factorial,

$$2^{2n}(n!)^2 > (2n)!$$

for all positive integers n and hence

$$\frac{x^{2n}}{2^{2n}(n!)^2} < \frac{x^{2n}}{(2n)!} \tag{66}$$

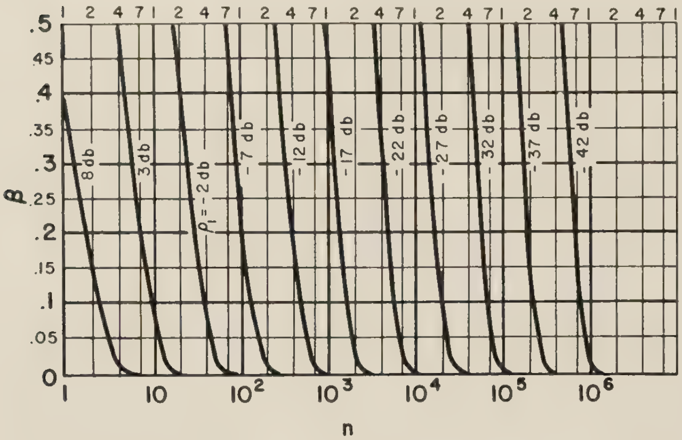


Fig. 4.

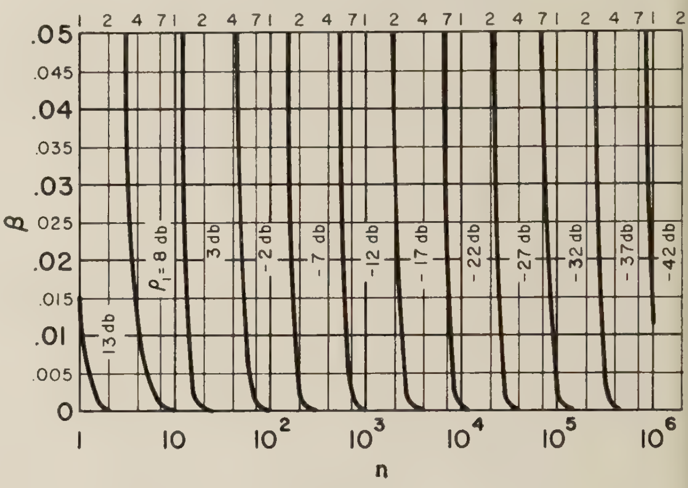


Fig. 5.

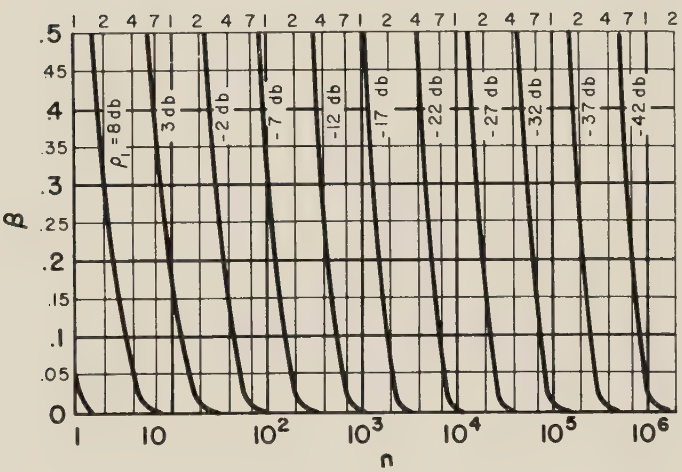


Fig. 6.

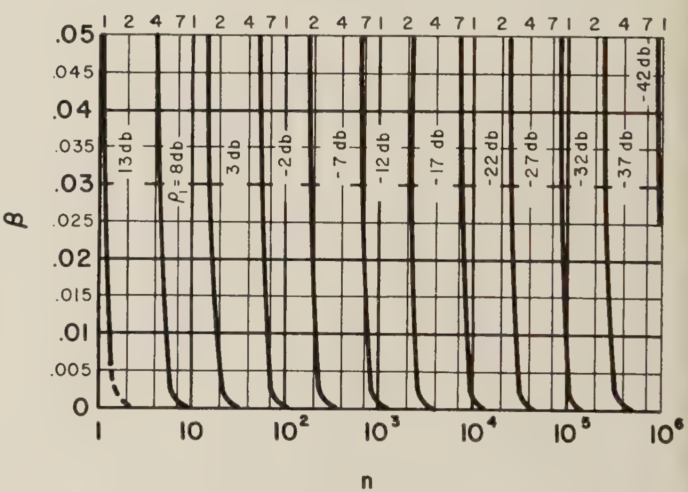


Fig. 7.

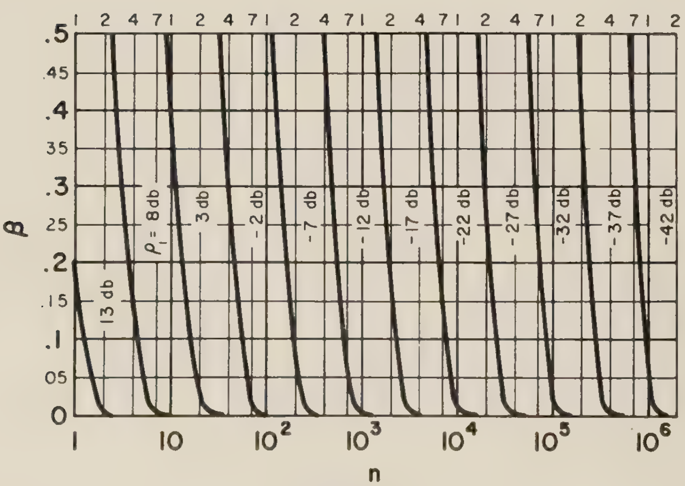


Fig. 8.

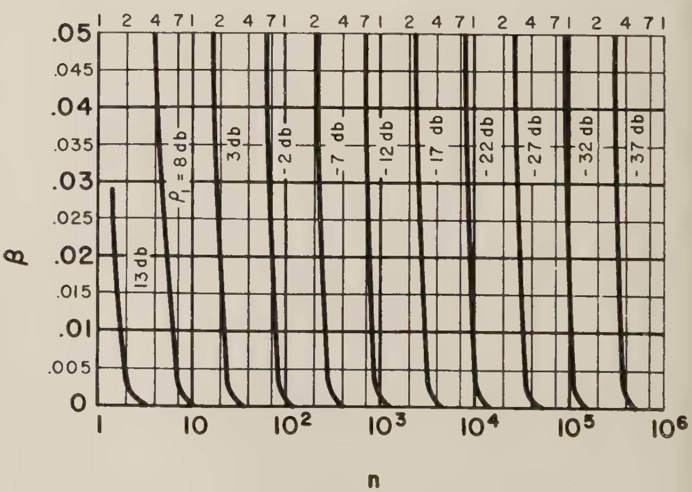


Fig. 9.

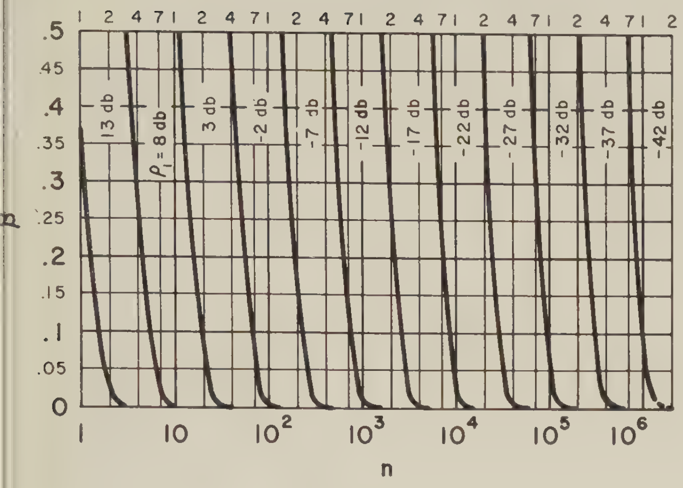


Fig. 10.

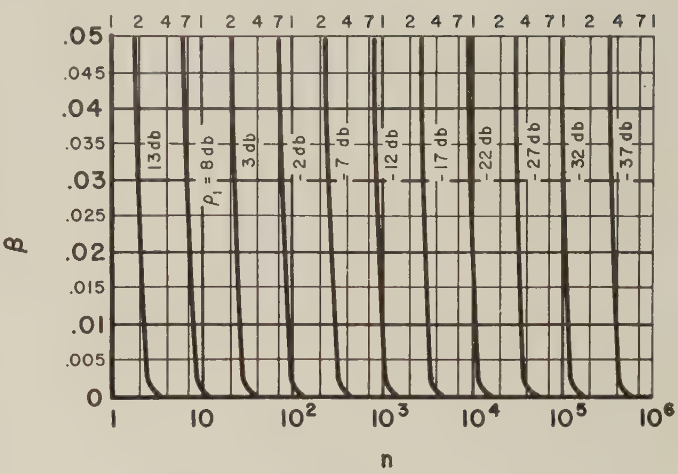


Fig. 11.

Figs. 4-11—Missed-signal probability β vs number of integration channels n with initial signal-to-noise ratio ρ_1 as parameter.

Fig.	Over-All False Alarm Probability	Missed-Signal Probability
4	$\alpha = 0.05$	$\beta \leqq 0.5.$
5	$\alpha = 0.05$	$\beta \leqq 0.05.$
6	$\alpha = 0.01$	$\beta \leqq 0.5.$
7	$\alpha = 0.01$	$\beta \leqq 0.05.$
8	$\alpha = 0.001$	$\beta \leqq 0.5.$
9	$\alpha = 0.001$	$\beta \leqq 0.05.$
10	$\alpha = 0.0001$	$\beta \leqq 0.5.$
11	$\alpha = 0.0001$	$\beta \leqq 0.05.$

for all nonzero x. The Bessel function $I_0(x)$ is defined by

$$I_0(x) = \sum_{n=0}^{\infty} \frac{1}{(n!)^2} \left(\frac{x}{2}\right)^{2n}$$

and using the inequality of (66)

$$I_0(x) < \sum_{n=0}^{\infty} \frac{x^{2n}}{(2n)!} < e^x, \quad \text{for} \quad x > 0.$$

[Of course $I_0(x)$ is asymptotic to $(1/\sqrt{2\pi x})\, e^x$; but in the present investigations, strict inequalities are desired.] Thus one may write (64) as

$$0 \leqq \beta(n) = \int_0^{\xi} ue^{-(u^2+b^2)/2} I_0(bu)\, du < \int_0^{\xi} ue^{-(u^2+b^2)/2} e^{bu}\, du$$

$$= \int_0^{\xi} ue^{-(u-b)^2/2}\, du. \tag{67}$$

It is now necessary to obtain an inequality on ξ. From the binomial theorem the inequality

$$\left(1 - \frac{x}{n}\right) \geqq (1 - x)^{1/n}, \qquad 0 \leqq x \leqq 1 \tag{68}$$

is readily deduced. Letting $x = 1 - a$ and recalling that $0 < a < 1$, the above inequality becomes

$$1 - a^{1/n} \geqq \frac{1 - a}{n}.$$

Thus

$$-\log (1 - a^{1/n}) \leqq -\log (1 - a) + \log n. \tag{69}$$

Since for n sufficiently large, $\log n > -\log (1 - a)$ one may write (69) as $-\log (1 - a^{1/n}) < 2 \log n$. Hence

$$\xi = \sqrt{-2 \log (1 - a^{1/n})} < 2\sqrt{\log n},$$

for n sufficiently large and (68) becomes

$$\beta(n) < \int_0^{2\sqrt{\log n}} ue^{-(u-b)^2/2}\, du. \tag{70}$$

Now $e^{-\frac{1}{2}(u-b)^2}$ will assume its maximum value when $|b - u|$ is as small as possible. For n large b is always greater than u for $0 \leqq u \leqq 2\sqrt{\log n}$. Thus $|b - u| > |b - 2\sqrt{\log n}|$ for n large and from (70)

$$\beta(n) < \int_0^{2\sqrt{\log n}} 2\sqrt{\log n}\, e^{-\frac{1}{2}(b - 2\sqrt{\log n})^2}\, du$$

$$= 4(\log n)\, e^{-\frac{1}{2}(\sqrt{2Qn} - 2\sqrt{\log n})^2}.$$

But $\sqrt{2Qn} > 4\sqrt{\log n}$ for n sufficiently large. Hence

$$\beta(n) < (4 \log n)e^{-2 \log n} = \frac{4 \log n}{n^2} < \frac{4}{n}$$

for n sufficiently large, and clearly

$$\lim_{n \to \infty} \beta(n) = 0.$$

It is clear from Cases 1, 3, and 4 that $\beta(n)$ must have at least one maximum. To determine this maximum the derivative $d\beta/db$ will be calculated and set equal to zero. From (64),

$$\frac{d\beta}{db} = \int_0^{\xi} ue^{-(u^2+b^2)/2} \left[-bI_0(bu) + \frac{d}{db} I_0(bu)\right] du$$

$$+ \xi e^{-(\xi^2+b^2)/2} I_0(b\xi) \frac{d\xi}{db}. \tag{71}$$

From the identity $dI_0(bu)/db = uI_1(bu)$ between the Bessel functions I_0 and I_1, the second part of the integral of (71) becomes

$$K = \int_0^\xi ue^{-(u^2+b^2)/2} \frac{d}{db} I_0(bu)\, du = \int_0^\xi u^2 e^{-(u^2+b^2)/2} I_1(bu)\, du.$$

Upon integrating by parts and recalling the further identity $d\,[uI_1(bu)]/du = buI_0(bu)$, we obtain

$$K = -\xi e^{-(\xi^2+b^2)/2} I_1(b\xi) + b \int_0^\xi ue^{-(u^2+b^2)/2} I_0(bu)\, du.$$

Substituting this formula in (71) yields

$$\frac{d\beta}{db} = \xi e^{-(\xi^2+b^2)/2}\left[I_0(b\xi)\frac{d\xi}{db} - I_1(b\xi)\right]$$

and β has a stationary point at the value of n which satisfies the equation

$$\frac{d\xi}{db} = \frac{I_1(b\xi)}{I_0(b\xi)}. \tag{72}$$

Now

$$\frac{d\xi}{db} = \frac{2}{b\xi n}\left(\log\frac{1}{a}\right)\frac{a^{1/n}}{1 - a^{1/n}}$$

and thus (72) may be written as, Case 5,

$$b\xi\frac{I_1(b\xi)}{I_0(b\xi)} = \frac{2}{n}\left(\log\frac{1}{a}\right)\frac{a^{1/n}}{1 - a^{1/n}}. \tag{73}$$

It is not hard to see that

$$0 < \frac{2}{n}\left(\log\frac{1}{a}\right)\frac{a^{1/n}}{1 - a^{1/n}} < 2$$

for all a with $0 < a < 1$ and all $n > 0$. Also

$$\frac{2}{n}\left(\log\frac{1}{a}\right)\frac{a^{1/n}}{1 - a^{1/n}}$$

is asymptotic to 2 as n approaches infinity while

$$b\xi\frac{I_1(b\xi)}{I_0(b\xi)} \sim b\xi$$

as n approaches infinity since $I_1(b\xi)/I_0(b\xi)$ is asymptotic to one. From this discussion of the functions appearing in (73) and their convexity (which is most conveniently investigated graphically but can also be done analytically) it can be concluded that (73) has a simple root, say at $n = n_0$. Thus β has a single maximum at $n = n_0$.

To summarize, the missed-signal probability β is asymptotic to zero as the number of channels n increases without limit and has a maximum at $n = n_0$ where n_0 is the only root of the scalar equation (73).

Acknowledgment

The authors would like to express their appreciation to L. E. Blumenson of the Electronics Research Laboratories for checking the derivations of the many complicated formulas, and to Miss J. Psygoda, also of the Laboratories, for performing the calculations necessary to construct the graphs of Figs. 4–11.

The Sequential Detection of a Sine-Wave Carrier of Arbitrary Duty Ratio in Gaussian Noise*

H. BLASBALG†

Summary—In this paper the Wald theory of sequential analysis is applied to the detection of a sine-wave carrier of arbitrary duty ratio in Gaussian noise. This is a generalization of a familiar problem. The detector law for the problem is obtained. In particular, it is specialized to the important cases: 1) arbitrary duty ratio and signal-to-noise ratio less than unity and 2) duty ratio much less than unity and peak-signal-to-noise ratio much greater than unity. For the latter case, it is shown that the best detector law goes over into a Bernoulli detector. In the former case it is shown that the only important parameter in detection is the average signal-power to noise-power ratio. For the case of unity duty ratio the detector law goes over into the familiar log $I_0(\eta r)$ characteristic. Furthermore, for threshold signals, it is shown that, in general, a first-order approximation to the logarithm of the likelihood ratio (or to the detector law) does not permit the sequential test to converge at one of the threshold parameters. A second-order approximation is always required. Curves of the operating characteristic function and the average sample number function are given for threshold signals.

* Manuscript received by the PGIT, May 10, 1957. This research was supported by the USAF through the Office of Sci. Res. of the Air Res. and Dev. Command. This paper represents a portion of the Doctor of Engineering Dissertation presented to The Johns Hopkins University, Baltimore, Md., May, 1956.

† Electronic Communications, Inc., Baltimore, Md. Formerly with The Johns Hopkins University Radiation Lab., Baltimore, Md.

I. Introduction

In this paper the detection of a pulsed carrier of arbitrary duty ratio in Gaussian noise is considered. To be more specific, we consider a set of observations at the independent time intervals $\Delta t_1, \Delta t_2, \cdots$ such that when a pulsed carrier of duty ratio d is transmitted, on the average, d per cent of the intervals are occupied by samples of carrier-plus-noise and $1 - d$ are occupied by noise only. In statistical language, d is the probability of measuring a signal-plus-noise sample in an arbitrary time interval on condition that a pulsed carrier is transmitted.

In the absence of a pulsed carrier all the intervals are occupied by noise. This corresponds to the case $d = 0$ and signal-to-noise ratio $\eta = 0$. We wish to find the sequential filter law for detecting the presence or absence of a pulsed carrier of arbitrary duty ratio $0 < d \leq 1$ and peak signal-to-noise ratio $\eta > 0$, in Gaussian noise. It is also desirable to investigate the detector law for those ranges of duty ratio and signal-to-noise ratio which are of practical interest.

An obvious application of this problem is the detection of pulse system activity in different regions of the spectrum. An electronically tuned receiver can be used to search the spectrum for pulse signals. Everything being equal, the more pulse signals that will appear in agiven frequency band the faster signal activity will be indicated. Hence, more time will be spent in those bands where the signal density is low and less time will be spent in the high density regions.

Another application of this detector is to the detection of targets in space by a high-speed search radar. In this application, the target detection time will decrease with the density of targets at a given range or the higher the density of targets the greater the range at which detection can be realized. This detector automatically trades duty ratio and signal-to-noise ratio for observation time in an efficient and natural way. It is certain that other applications exist.

II. Brief Formulation of Sequential Detection Theory

A sequential test [1], [8] can be described as follows: at $t = t_1$ an observer measures the sample value r_1. Based on this value he must decide whether to accept the hypothesis H_0 that the parameter of the distribution is $\eta \leq \eta_0$ or H_1 that $\eta \geq \eta_1$, $(\eta_0 < \eta_1)$ or he must decide that the datum is insufficient to accept either one of the hypotheses with confidence. If H_0 or H_1 is accepted, the experiment is terminated. On the other hand if the datum is insufficient to lead to the acceptance of one of these hypotheses, then at $t = t_2$ the observer takes another sample r_2. Based on the sample of size two (r_1, r_2) the observer must once again make one of three possible decisions: accept H_0 or H_1 or the datum is insufficient for accepting either. If the hypothesis H_0 that only noise is present or H_1 that signal-plus-noise is present is accepted, the experiment is terminated. If the data are insufficient for a decision, r_3 is observed. The same decision procedure is then repeated on the sample point (r_1, r_2, r_3). Experimentation is continued in this manner until either H_0 or H_1 is accepted. It is clear that the number of samples required for the termination of a sequential test with the acceptance of the hypothesis H_0 or H_1 is a random variable.

In a sequential test the observer specifies the error probability α of accepting the hypothesis H_1 when H_0 is true and the error probability β of accepting H_0 when H_1 is true. From the class of all sequential tests, the decision rule which minimizes the average number of samples required for termination at the threshold parameters $\eta = \eta_0$ and $\eta = \eta_1$ is chosen. The rule for independent sampling is given by the following: accept H_0 that noise is present, at that value of the sample number n for which,

$$\sum_{i=1}^{n} \log \frac{P(r_i, \eta_1)}{P(r_i, \eta_0)} \leq \log \frac{\beta}{1-\alpha} \rightarrow H_0, \tag{1}$$

and accept H_1 that signal-plus-noise is present at that value of n for which,

$$\sum_{i=1}^{n} \log \frac{P(r_i, \eta_1)}{P(r_i, \eta_0)} \geq \log \frac{1-\beta}{\alpha} \rightarrow H_1. \tag{2}$$

$P(r, \eta)$ = probability density function of the random variable r when η is the true parameter,
α = probability of accepting H_1 when H_0 is true,
$1 - \alpha$ = probability of accepting H_0 when H_0 is true,
β = probability of accepting H_0 when H_1 is true,
$1 - \beta$ = probability of accepting H_1 when H_1 is true.

The left side of Eq. (1) or (2) is the logarithm of the probability ratio or likelihood ratio.

The two most important characteristics for judging the performance of a sequential detector are the operating characteristic function (oc function) and the average sample number function (asn function). The oc function $L(\eta)$ gives the probability of accepting the hypothesis H_0 as a function of the parameter η of the distribution function. It expresses the confidence in the decision.

The mathematical expression for this function is given by the pair of parametric equations,

$$L_\eta(h) = \frac{\left(\frac{1-\beta}{\alpha}\right)^h - 1}{\left(\frac{1-\beta}{\alpha}\right)^h - \left(\frac{\beta}{1-\alpha}\right)^h}; \qquad -\infty \leq h \leq \infty \tag{3}$$

and,

$$E_\eta(e^{hz}) = 1, \tag{4}$$

where,

$$z = \log \frac{P(r, \eta_1)}{P(r, \eta_0)}. \tag{5}$$

$E_\eta[e^{hz}]$ = expected value of random variable e^{hz}.

For a given α, β, $L_q(h)$ is a universal function of h. The parameter η of the distribution is related to h by (4). From (3) and (4) the oc function $L(\eta)$ can be obtained.

The asn function gives the average number of samples required for the termination of a sequential test as a function of the parameter η. It is an expression of the cost of experimentation in terms of the number of samples. The mathematical expression for the asn function is given by,

$$E_\eta(n) = \frac{L(\eta) \log \frac{\beta}{1-\alpha} + [1 - L(\eta)] \log \frac{1-\beta}{\alpha}}{E_\eta(z)}. \tag{6}$$

At the indeterminate point $h = 0$ corresponding to a value of $\eta = \eta'$ at which $E_\eta(z) = 0$, we have for the oc function,

$$L(\eta') = \frac{\log \frac{1-\beta}{\alpha}}{\log \frac{1-\beta}{\alpha} + \log \frac{1-\alpha}{\beta}} \tag{7}$$

and for the asn function,

$$E_{\eta'}(n) = \frac{\left(\log \frac{1-\beta}{\alpha}\right)\left(\log \frac{1-\alpha}{\beta}\right)}{E_{\eta}'(z^2)}. \tag{8}$$

The oc and asn functions are approximations since the excess over the boundaries of (1) and (2) is neglected. The approximations are however extremely good for practical applications [8]. The results of sequential theory briefly stated will now be applied to the detection of a pulsed sine-wave-carrier of arbitrary duty ratio in Gaussian noise.

III. The Detection of a Sine-Wave Carrier of Arbitrary Duty Ratio in Gaussian Noise

The probability density function of the envelope of Gaussian noise after passing through a narrow band-pass filter is given by [5]

$$P_0(r) = re^{-\frac{1}{2}r^2} \tag{9}$$

where,

$r = R/\sigma_N$ = normalized random variable,
R = envelope voltage sample,
σ_N = rms value of noise at input to band-pass filter.

The probability density function of the envelope of a sine-wave carrier plus noise is [5],

$$P_\eta(r) = re^{-\frac{1}{2}(r^2+\eta^2)}I_0(\eta r) \tag{10}$$

where,

$\eta = \frac{A}{\sigma_N}$ = peak-signal-to-rms noise ratio,

$I_0(\eta r) = \sum_{k=0}^{\infty} \left(\frac{\eta r}{2}\right)^{2k} \frac{1}{(k!)^2}$ = zero order modified Bessel function.

When $\eta = 0$, (10) reduces to (9).

Let d be the duty ratio of the pulsed carrier in the sense described in Section I. In the absence of a carrier only noise is present. Hence, the probability density is given by (9). The probability under H_1 of the random variable r exceeding some arbitrary threshold is equal to the probability of occurrence of signal-plus-noise in any one of the intervals times the probability that signal-plus-noise exceeds the threshold plus the probability of observing noise times the probability that the noise exceeds the threshold. The distribution function of a pulsed carrier in noise is therefore,

$$P_{\eta,d}(r) = dP_\eta(r) + [1-d]P_0(r), \qquad 0 < d \leq 1 \tag{11}$$

where $P_0(r)$ and $P_\eta(r)$ are given by (9) and (10). From (11) it is seen that $d = 0$, $\eta = 0$ simultaneously, or $d = 0$, or $\eta = 0$, all lead to the distribution $P_0(r)$. This is consistent with the manner in which the problem is formulated.

In order to detect the presence or absence of a carrier in noise, we test the hypothesis H_0 that the signal-to-noise ratio is $\eta = 0$ (implying $d = 0$) against the alternative hypothesis H_1 that $\eta > 0$ (implying $d > 0$). Let $\eta_1 > 0$ and $d_0 > 0$ be some numbers arbitrarily close to zero. Then for practical reasons, we must test the hypothesis $\eta = 0$ against $\eta \geq \eta_1$. This also implies $d = 0, d \geq d_0$. In order to obtain the decision regions, we make use of the general results given by (1) and (2). The logarithm of the probability ratio corresponding to a particular sample r_i is obtained from (11) as,

$$z_i = \log \frac{P_{\eta_1,d_0}(r_i)}{P_0(r_i)} = \log \left[1 + d_0\left(\frac{P_{\eta_1}(r_i)}{P_0(r_i)} - 1\right)\right]. \tag{12}$$

From (9) and (10), the ratio

$$\frac{P_{\eta_1}(r_i)}{P_0(r_i)} = e^{-\frac{1}{2}\eta_1^2}I_0(\eta_1 r_i). \tag{13}$$

Corresponding to an observation of n independent samples we have from (1), (2), (12), and (13) the decision regions,

$$\sum_{i=1}^{n} \log [1 + d_0(e^{-\frac{1}{2}\eta_1^2}I_0(\eta_1 r_i) - 1)] \leq \log \frac{\beta}{1-\alpha} \rightarrow H_0 \tag{14}$$

and,

$$\sum_{i=1}^{n} \log [1 + d_0(e^{-\frac{1}{2}\eta_1^2}I_0(\eta_1 r_i) - 1)] \geq \log \frac{1-\beta}{\alpha} \rightarrow H_1. \tag{15}$$

The arrow indicates the hypothesis to which the decision region belongs. For the special case, $d = 1$, which corresponds in Section I to the case where all the observation intervals are occupied by signal-plus-noise (when signal is in fact present) the decision regions reduce to the familiar form,

$$\sum_{i=1}^{n} \log I_0(\eta_1 r_i) \leq \log \frac{\beta}{1-\alpha} + \frac{n\eta_1^2}{2} \rightarrow H_0 \tag{16}$$

and

$$\sum_{i=1}^{n} \log I_0(\eta_1 r_i) \geq \log \frac{1-\beta}{\alpha} + \frac{n\eta_1^2}{2} \rightarrow H_1. \tag{17}$$

The detector for the special case above is the familiar $\log I_0(\eta_1 r)$ detector. This result has been obtained when considering the detection of a target at a particular range by means of a pulsed radar [6]. The detector law given by (14) and (15) for the more general case is a complicated computer. Furthermore, machine methods are required to calculate the oc and asn functions. We will therefore specialize the results to those ranges of the parameters d_0 and η_1 that are of practical interest.

IV. Sequential Detector Law and Characteristics For a Pulsed Carrier of Arbitrary Duty Ratio in Noise of Signal-To-Noise Ratio $\eta_1 < 1$

The case now considered is the detection of threshold signals in noise, or signals of arbitrary duty ratio and signal-to-noise ratio less than unity. Consider a particular term in (14) or (15),

$$z = \log\left[1 + d_0(e^{-\frac{1}{2}\eta_1^2} I_0(\eta_1 r) - 1)\right]. \tag{18}$$

The first two terms in the Taylor series approximation to the logarithm of (18) are given by,

$$z \approx d_0[e^{-\frac{1}{2}\eta_1^2} I_0(\eta_1 r) - 1] - \frac{d_0^2}{2}[e^{-\frac{1}{2}\eta_1^2} I_0(\eta_1 r) - 1]^2. \tag{19}$$

Let us further expand (19) in a Taylor series in powers of η_1 by using the expansion for $I_0(\eta_1 r)$. Then, neglecting all orders higher than the fourth leads to

$$\begin{aligned} z(r; d_0, \eta_1) = &-\frac{d_0\eta_1^2}{2}\left[1 - \frac{\eta_1^2}{4}(1 - d_0)\right] \\ &+ \frac{d_0\eta_1^2 r^2}{4}\left[1 - \frac{\eta_1^2}{2}(1 - d_0)\right] \\ &+ \frac{d_0\eta_1^4 r^4}{32}\left[\frac{1}{2} - d_0\right]. \end{aligned} \tag{20}$$

Eq. (20) gives an approximation to the detector law for the case where the signal-to-noise ratio is $\eta_1 < 1$ and for arbitrary duty ratio. Eq. (20) shows that the detector law contains a bias given by the first term and a linear combination of second and fourth order terms. The reason for including the fourth order term will become evident shortly. For the special case, $d_0 = 1$, we have,

$$z(r; 1, \eta_1) = -\frac{\eta_1^2}{2} + \frac{\eta_1^2 r^2}{4} - \frac{\eta_1^4 r^4}{64}. \tag{21}$$

This expression is equivalent to that obtained in [4] and serves as somewhat of a check on the validity of (20).

Let $E_{\eta,d}(z)$ be the expected value of $z(r; d_0, \eta_1)$ on condition that a carrier of signal-to-noise ratio η and duty ratio d is present or

$$E_{\eta,d}(z) = \int z(r; d_0, \eta_1) P_{\eta,d}(r)\, dr, \tag{22}$$

where $P_{\eta,d}(r)$ is defined by (11) in conjunction with (9) and (10). From [7] we know that,

$$E_\eta(r^2) = \int_0^\infty r^2 P_\eta(r)\, dr = \eta^2 + 2, \tag{23}$$

where $P_\eta(r)$ is given by (10) and,

$$E_\eta(r^4) = \int_0^\infty r^4 P_\eta(r)\, dr = 8 + 8\eta^2 + \eta^4. \tag{24}$$

From (11) and using the above results we have,

$$\begin{aligned} E_{\eta,d}(r^2) &= d(2 + \eta^2) + 2(1 - d), \\ &= 2 + d\eta^2 \end{aligned} \tag{25}$$

and,

$$\begin{aligned} E_{\eta,d}(r^4) &= d(8 + 8\eta^2 + \eta^4) + 8(1 - d) \\ &\quad 8 + 8d\eta^2 + d\eta^4. \end{aligned} \tag{26}$$

Making use of (25) and (26) and applying these to (22) yields

$$E_{\eta,d}(z) = \frac{d\, d_0 \eta^2 \eta_1^2}{4} - \frac{d_0^2 \eta_1^4}{8} \qquad \eta\eta_1 < 1. \tag{27}$$

In this development certain terms have been neglected. However, (27) is a good approximation for the range of values indicated and in particular in the interval $0 \le \eta \le \eta_1$. Eq. (27) gives the average value of $z(r; d_0, \eta_1)$. When $d = d_0 = 1$ and $\eta = \eta_1$ we obtain the result equivalent to that obtained in [4],

$$E_{\eta_1}(z) = \frac{\eta_1^4}{8}. \tag{28}$$

Similarly, when $\eta = 0$ we obtain,

$$E_0(z) = -\frac{\eta_1^4}{8}. \tag{29}$$

This once again is a check on the expansions.

Suppose we consider only the first term in the expansion of the logarithm of the likelihood ratio given in (19).

$$z_0 \approx d_0[e^{-\frac{1}{2}\eta_1^2} I_0(\eta_1 r) - 1]. \tag{30}$$

If we take the average value of z_0 on condition that only noise is present we have,

$$\begin{aligned} E_0(z_0) = d_0 &\int_0^\infty r e^{-\frac{1}{2}(\eta_1^2 + r^2)} I_0(\eta_1 r)\, dr \\ &- d_0 \int_0^\infty r e^{-\frac{1}{2}r^2}\, dr = 0, \end{aligned} \tag{31}$$

since the integrals represent the areas under the given probability density functions. The asn function given in (6) contains the parameter $E_\eta(z)$ in the denominator. For the approximation considered in (30), $E_n(z_0) = 0$ at $\eta = 0$. Hence, the asn function is infinite at this point or the sequential test does not converge in probability when noise is present. We therefore conclude that at least two terms in the approximation to the logarithm of the likelihood ratio are necessary for the sequential test to converge. This requires that for small signal-to-noise ratios the detector law must include a fourth order term as well as second order and constant terms. This property was also observed in [4]. We will later prove, in general, that a sequential probability ratio test never converges at the threshold parameter $\theta = \theta_0$ when a first order approximation to the logarithm of the likelihood ratio is used.

The statistical decision regions for the acceptance of the hypothesis H_0 that noise is present or H_1 that signal-plus-noise is present is obtained from (1), (2), and (20) as,

$$\sum_{i=1}^{n} z(r_i; d_0, \eta_1) \le \log \frac{\beta}{1 - \alpha} \to H_0, \tag{32}$$

$$\sum_{i=1}^{n} z(r_i; d_0, \eta_1) \geq \log \frac{1-\beta}{\alpha} \rightarrow H_1. \tag{33}$$

V. The Operating Characteristic Function (OC Function) for $\eta_1 < 1$ and Arbitrary Duty Ratio

The oc function is given by the parametric (3) and (4) or,

$$L_\eta(h) = \frac{\left(\frac{1-\beta}{\alpha}\right)^h - 1}{\left(\frac{1-\beta}{\alpha}\right)^h - \left(\frac{\beta}{1-\alpha}\right)^h}; \quad -\infty \leq h \leq \infty \tag{34}$$

and,

$$E_\eta(e^{zh}) = 1, \tag{35}$$

where z is given by (20). Let us expand (35) in a Taylor series and neglect all orders higher than the second. For the problem considered we have,

$$h(\eta, d) = -2 \frac{E_{\eta,d}(z)}{E_{\eta,d}(z^2)}. \tag{36}$$

The numerator of (36) is given by (27). The denominator can be obtained from (20) with the help of relationships (25) and (26). Neglecting orders in η_1 higher than the fourth, we have,

$$h(\eta, d) = \frac{-2 \frac{d}{d_0}\left(\frac{\eta}{\eta_1}\right)^2 + 1}{1 + d\eta^2}. \qquad \eta\eta_1 < 1. \tag{37}$$

In the region $0 \leq \eta \leq \eta_1$, which is actually the region of interest, we have,

$$h(\eta, d) = -2 \frac{d}{d_0}\left(\frac{\eta}{\eta_1}\right)^2 + 1. \qquad 0 \leq \eta \leq \eta_1. \tag{38}$$

(This is a good approximation when $\eta_1 < 0.3$).

At the threshold parameters $\eta = 0$ we have $h = 1$ and when $d = d_0$, $\eta = \eta_1$ we have $h = -1$, as required. This serves as a partial check of the approximation (38) since for all sequential tests h takes on these values at the threshold parameters. When $\eta = \eta_1$ and d is some arbitrary number we have,

$$h(\eta_1, d) = -2 \frac{d}{d_0} + 1. \tag{39}$$

Substituting (38) into (34) we obtain an approximation to the oc function as,

$$L(\eta, d) = \frac{\left(\frac{1-\beta}{\alpha}\right)^{[-2(d/d_0)(\eta/\eta_1)^2+1]} - 1}{\left(\frac{1-\beta}{\alpha}\right)^{[-2(d/d_0)(\eta/\eta_1)^2+1]} - \left(\frac{\beta}{1-\alpha}\right)^{[-2(d/d_0)(\eta/\eta_1)^2+1]}}; \quad 0 \leq \eta \leq 0.3. \tag{40}$$

We now define the following:

$P_0 = d_0\eta_1^2$ = minimum detectable average signal-power-to-noise power ratio,

$P = d\eta^2$ = average signal-power-to-noise-power ratio actually received.

Then (40) becomes,

$$L(P) = \frac{\left(\frac{1-\beta}{\alpha}\right)^{[-2(P/P_0)+1]} - 1}{\left(\frac{1-\beta}{\alpha}\right)^{[-2(P/P_0)+1]} - \left(\frac{\beta}{1-\alpha}\right)^{[-2(P/P_0)+1]}}. \tag{41}$$

It is therefore seen that when the peak-signal-to-rms-noise ratio is less than unity the error probabilities given by the oc function depend only on the average power. At the indeterminate points $P' = P_0/2$ or $(d\eta)' = d_0\,\eta_1^2/2$ we have from (7)

$$L(P') = \frac{\log \frac{1-\beta}{\alpha}}{\log \frac{1-\beta}{\alpha} + \log \frac{1-\alpha}{\beta}}. \tag{42}$$

Fig. 1 is a curve of the oc function for $\alpha = \beta = 0.01$ and minimum detectable signal-power-to-noise-power ratios $P_0 = 0.01, 0.10$.

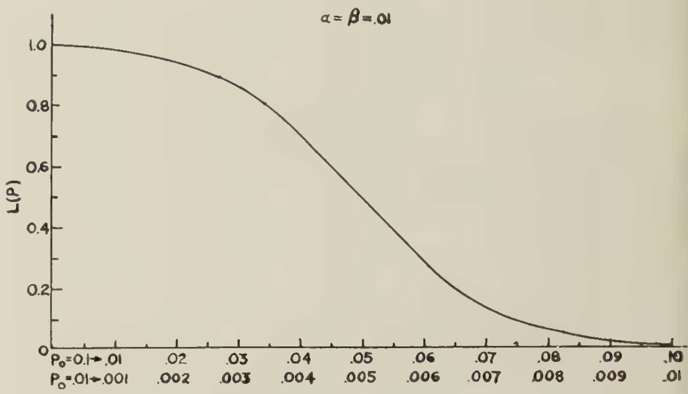


Fig. 1—OC function for sinusoidal carrier of arbitrary duty ratio in Gaussian noise of signal-to-noise ratio, $\eta \ll 1$.

VI. The Average Sample Number (ASN) Function for $\eta_1 < 1$

The general expression for the asn function is given by (6). When specialized to the problem considered we have from (27) and (6).

$$E_{\eta,d}(n) = \frac{L(\eta, d) \log \frac{\beta}{1-\alpha} + [1 - L(\eta, d)] \log \frac{1-\beta}{\alpha}}{\frac{d\, d_0 \eta^2 \eta_1^2}{4} - \frac{d_0^2 \eta_1^4}{8}}, \tag{43}$$

where $L(\eta, d)$ is the oc function given in (40). In terms of the parameter P previously defined we have,

$$E_P(n) = \frac{L(P) \log \frac{\beta}{1-\alpha} + [1 - L(P)] \log \frac{1-\beta}{\alpha}}{\frac{PP_0}{4} - \frac{P_0^2}{8}}. \tag{44}$$

At the indeterminate point $P' = P_0/2$ we obtain from (8)

$$E_{P'}(n) = \frac{\left(\log \frac{1-\beta}{\alpha}\right)\left(\log \frac{1-\alpha}{\beta}\right)}{\frac{P_0^2}{4}}. \tag{45}$$

At the threshold parameters $P = 0$, $P = P_0$ it is seen that the asn function is inversely proportional to the average signal-power-to-noise-power ratio squared or the fourth power of the signal-to-noise ratio. Fig. 2 is a curve of the asn function for $\alpha = \beta = 0.01$ and two values of threshold average power $P_0 = 0.01, 0.10$.

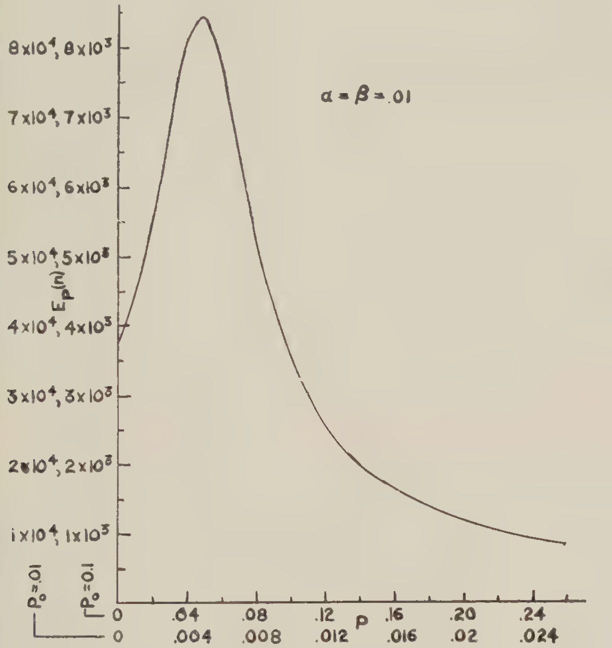


Fig. 2—ASN function for sinusoidal carrier of arbitrary duty ratio in Gaussian noise of signal-to-noise ratio, $\eta \ll 1$.

VII. To Show that a Second Order Approximation to The Logarithm of the Probability Ratio is Required for a Sequential Test to Converge at the Threshold Parameter, $\theta = \theta_0$

Let

$$z = \log \frac{P(r, \theta_1)}{P(r, \theta_0)}, \tag{46}$$

be a particular term in the logarithm of the likelihood ratio corresponding to a test of hypothesis $\theta \leq \theta_0$ against $\theta \geq \theta_1$, $(\theta_1 > \theta_0)$. The functions $P(r, \theta_1)$, $P(r, \theta_0)$ are probability density functions belonging to a one parameter family of density functions. Expanding z in the familiar Taylor series for the logarithm yields,

$$z \approx \left(\frac{P(r, \theta_1)}{P(r, \theta_0)} - 1\right) - \frac{1}{2}\left(\frac{P(r, \theta_1)}{P(r, \theta_0)} - 1\right)^2 + \frac{1}{3}\left(\frac{P(r, \theta_1)}{P(r, \theta_0)} - 1\right)^3 - \cdots. \tag{47}$$

The expected value of z on condition that θ_0 is the true parameter upon neglecting third order terms in z is simply,

$$E_{\theta_0}(z) = -\frac{1}{2}\int_{-\infty}^{\infty} \frac{[P(r, \theta_1) - P(r, \theta_0)]^2}{P(r, \theta_0)}\, dr. \tag{48}$$

The first term in the expression (47) contributes nothing to $E_{\theta_0}(z)$, since,

$$\int_{-\infty}^{\infty} P(r, \theta_1)\, dr = \int_{-\infty}^{\infty} P(r, \theta_0)\, dr = 1. \tag{49}$$

Since the asn function, (6), contains $E_{\theta_0}(z)$ in the denominator it follows that at the point $\theta = \theta_0$, $E_{\theta_0}(n) = \infty$, if only the first term,

$$z^{(1)} \approx \frac{P(r, \theta_1)}{P(r, \theta_0)} - 1, \tag{50}$$

is used in the approximation (47). At the parameter point $\theta = \theta_1$ the expected value of (50) does not vanish.

Although the asn function is an approximation since the excess over the boundaries is neglected we claim that the proof is still valid. Neglecting the excess over the boundaries at the termination of a sequential test introduces an approximation in the numerator of the asn function only and not in the denominator. The expression $E_\eta(z)$ in the denominator is exact and it is shown here that a first order approximation to z yields $E_\eta(z) = 0$ at $\eta = 0$. This is actually the cause of the divergence of the sequential test at $\eta = 0$. Furthermore, when $\eta_1 \ll 1$, the excess over the boundaries at the termination of the experiment is negligible and hence the expression for the asn function is almost exact.

As an example consider the detector law given in (16)

$$z = \log I_0(\eta_1 r) - \tfrac{1}{2}\eta_1^2. \tag{51}$$

A first order approximation to $\log I_0(\eta_1 r)$ is given by

$$\log I_0(\eta_1 r) \approx I_0(\eta_1 r) - 1. \tag{52}$$

Using the series for $I_0(\eta_1 r)$ given in (10), substituting into (52), and considering only the first nonvanishing term, gives,

$$\log I_0(\eta_1 r) \approx \frac{\eta_1^2 r^2}{4}. \tag{53}$$

Hence the detector for the case $\eta_1 < 1$ appears to be

$$z^{(1)} = \frac{\eta_1^2 r^2}{4} - \tfrac{1}{2}\eta_1^2. \tag{54}$$

This is the so-called biased square-law detector. However, if we take the expected value of (54) on condition that only noise is present, we have,

$$E_0(z^{(1)}) = 0$$

since from (23) $E_0(r^2) = 2$. However, at the threshold parameter $\eta = \eta_1$, using (23), we have

$$E_{\eta_1}(z^{(1)}) = \frac{\eta_1^4}{4}. \tag{55}$$

We can therefore conclude that the square law detector plus a bias term contributes the required information for a terminal decision when signal is present, but it does not contribute any information, on the average, when noise alone is present. From (21) it is seen that when noise alone is present the fourth order term contributes the required information $E_0(z)$ for a terminal decision. Comparing (55) to (28), it is seen that when signal is present the average number of samples required for termination for the required detector, (21), is twice as large when compared to the detector given by (54). Thus, the fact that termination is required when $\eta = 0$ as well as when $\eta = \eta_1$, leads to an increase in the number of samples when signal is present. Therefore, we gain convergence at $\eta = 0$ at the expense of losing one half the information at $\eta = \eta_1$. The amount of information lost is that required for the sequential test to terminate at $\eta = 0$. [Here we interpret information as the quantity $E_\theta(z)$.]

VIII. To Show That for $\eta_1 > 1$ and $d_0 \ll 1$ the Optimum Detector Law is a Bernoulli Detector

Let us refer once again to the expression for the general detector law for n samples,

$$z(n) = \sum_{i=1}^{n} \log\left[1 + d_0(e^{-\frac{1}{2}\eta_1^2} I_0(\eta_1 r_i) - 1)\right]. \tag{56}$$

For $\eta_1 r > 2$ we can approximate [5] $I_0(\eta_1 r)$ by the asymptotic expression,

$$I_0(\eta_1 r) \approx \frac{e^{\eta_1 r}}{\sqrt{2\pi\eta_1 r}}. \tag{57}$$

It can be shown that when $\eta_1 = 2$, the probability that r exceeds unity on condition $\eta_1 = 2$ is greater than 0.90 or expressed mathematically,

$$p_{\eta_1}(r \geq 1) \geq 0.90; \qquad \eta_1 \geq 2.$$

Hence, the approximation is valid almost always. Putting (57) into (56) gives for the detector law,

$$z(n) \approx \sum_{i=1}^{n} \log\left[1 + d_0\left(\frac{e^{[-\frac{1}{2}\eta_1^2 + \eta_1 r_i]}}{\sqrt{2\pi\eta_1 r_i}} - 1\right)\right]. \tag{58}$$

If we further approximate the logarithm by its first two terms in the Taylor expansion, we have, after some manipulation,

$$z(n) = \sum_{i=1}^{n} d_0\left(\frac{e^{[-\frac{1}{2}\eta_1^2 + \eta_1 r_i]}}{\sqrt{2\pi\eta_1 r_i}} - 1\right) \cdot\left[1 - \frac{1}{2} d_0\left(\frac{e^{[-\frac{1}{2}\eta_1^2 + \eta_1 r_i]}}{\sqrt{2\pi\eta_1 r_i}} - 1\right)\right]. \tag{59}$$

Furthermore, assume that the parameters d_0 and η_1 are such that inequality,

$$\frac{d_0}{2}\left(\frac{e^{[-\frac{1}{2}\eta_1^2 + \eta_1 r]}}{\sqrt{2\pi\eta_1 r}} - 1\right) \ll 1, \tag{60}$$

is almost always satisfied. In order to obtain an upper bound on d_0 for which (60) holds, we consider the following approach. The standard deviation of r for values of $\eta_1 > 2$ is approximately,

$$\sigma_{\eta_1} = \sqrt{2 + \eta_1^2 - (\eta_1^2)} = \sqrt{2}. \tag{61}$$

This follows from the fact that the mean square value of r is $(2 + \eta_1^2)$ and the squared average value of r for $\eta_1 > 2$ is approximately η_1^2. The probability that r is less than $\eta_1 + 3\sqrt{2}$ (where $3\sqrt{2}$ = three standard deviations) is certainly greater than 0.90 when $\eta_1 \geq 2$ or expressed mathematically,

$$p_{\eta_1}\{r \leq \eta_1 + 3\sqrt{2}\} \geq 0.90. \tag{62}$$

Hence, replacing r in (60) by $\eta_1 + 3\sqrt{2}$ strengthens the inequality of (60) almost always. Therefore,

$$\frac{d_0}{2}\left(\frac{e^{-[\frac{1}{2}\eta_1^2 + 3\sqrt{2}\eta_1]}}{\sqrt{2\pi\eta_1}} - 1\right) \ll 1. \tag{63}$$

We replaced r in the denominator by η_1 strengthening the inequality even more. Neglecting -1 in the brackets, and taking logarithms of (63) gives

$$\log\frac{d_0}{2} \ll \log\sqrt{2\pi\eta_1} - (\tfrac{1}{2}\eta_1^2 + 3\sqrt{2}\eta_1),$$

$$\ll -(\tfrac{1}{2}\eta_1^2 + 3\sqrt{2}\eta_1)$$

or

$$d_0 \ll 2e^{-(\frac{1}{2}\eta_1^2 + 3\sqrt{2}\eta_1)}. \tag{64}$$

For a given value $\eta_1 \geq 2$ if d_0 is chosen so that inequality (64) is satisfied then inequality (60) is almost always satisfied. Hence, we can further approximate (59) by considering the first term only, or,

$$z(n) \approx d_0 e^{-\frac{1}{2}\eta_1^2} \sum_{i=1}^{n} \frac{e^{\eta_1 r_i}}{\sqrt{2\pi\eta_1 r_i}} - d_0 n. \tag{65}$$

The detector law of (65) weights the peaks of the signal very heavily.

Since $d_0 \ll 1$, it follows that on the average the number samples is $n \gg 1$, as many samples must be observed in order to allow for the occurrence of at least one pulse when signal is present. With these assumptions we approximate the sum in (65) by an integral or,

$$I = \frac{1}{n}\sum_{i=1}^{n} \frac{e^{\eta_1 r_i}}{\sqrt{2\pi\eta_1 r_i}} \approx \int_0^\infty re^{-\frac{1}{2}r^2} \frac{e^{\eta_1 r}}{\sqrt{2\pi\eta_1 r}}\, dr$$

$$= e^{\frac{1}{2}\eta_1^2}\int_0^\infty \frac{re^{-\frac{1}{2}(r-\eta_1)^2}}{\sqrt{2\pi\eta_1 r}}\, dr. \tag{66}$$

Again since most of the samples will be noise [*i.e.*, $(1 - d_0)n$ for large n] we average on condition that noise is present. Before integrating (66) we note that the integrand contributes to the integral primarily in the neighborhood of η_1. Hence, the term $1/\sqrt{r}$ has negligible effect on the

ntegral everywhere except at $r = \eta_1$. We further choose η_1 sufficiently large ($\eta_1 > 3$) and replace the range of integration to cover the entire real line. After all these assumptions, we finally obtain,

$$I \approx \frac{e^{\frac{1}{2}\eta_1^2}}{\eta_1} \int_{-\infty}^{\infty} \frac{re^{-\frac{1}{2}(r-\eta_1)^2}}{\sqrt{2\pi}}\, dr = e^{\frac{1}{2}\eta_1^2}. \tag{67}$$

Thus,

$$\sum_{i=1}^{n} \frac{e^{\eta_1 r_i}}{\sqrt{2\pi}\eta_1 r_i} \approx ne^{\frac{1}{2}\eta_1^2}. \tag{68}$$

Eq. (68) is, of course, only an approximation. The probability that a noise sample will exceed the threshold $\eta_1 - 3\sqrt{2}$ (*i.e.*, three standard deviations to the left of η_1) is

$$p_0(r) = e^{-\frac{1}{2}[\eta_1 - 3\sqrt{2}]^2} \approx e^{-\frac{1}{2}\eta_1^2} \tag{69}$$

for $\eta_1 \gg 3\sqrt{2}$.

The number of samples exceeding the threshold is approximately,

$$p_0(\eta_1)n = k_n. \tag{70}$$

Substituting (70) into (68) for n gives,

$$\sum_{i=1}^{n} \frac{e^{\eta_1 r_i}}{\sqrt{2\pi}\eta_1 r_i} \approx k_n e^{\eta_1^2}. \tag{71}$$

Hence, for $\eta_1 \gg 1$ and very small d_0, the sum in (71) contributes approximately the same information as a Bernoulli detector set at a slicing threshold $r_0 = \eta_1 - 3\sqrt{2}$ which weights a success by $e^{\eta_1^2}$. If we are given a sequence of independent random variables $r_1, \cdots r_n$, then a Bernoulli sequence can be derived from the given sequence by the following slicing operation:

$$F(r_i; r_0) = X_i,$$

where,

$$X_i = 1 \quad \text{when} \quad r_i \geq r_0$$
$$= 0 \quad \text{when} \quad r_i < r_0.$$

The sequence $\{X_i\}$ $i = 1, 2, \cdots n$ consists of zeros and ones distributed in a Bernoulli distribution. Substituting (71) into (65) gives,

$$z(n) = d_0 k_n e^{\frac{1}{2}\eta_1^2} - d_0 n. \tag{72}$$

The decision regions for detecting H_0 and H_1 are therefore,

$$k_n \leq \frac{1}{d_0 e^{\frac{1}{2}\eta_1^2}} \left[\log \frac{\beta}{1-\alpha} + d_0 n\right] \to H_0 \tag{73}$$

and,

$$k_n \geq \frac{1}{d_0 e^{\frac{1}{2}\eta_1^2}} \left[\log \frac{1-\beta}{\alpha} + d_0 n\right] \to H_1 \tag{74}$$

which holds when the conditions,

$$d_0 \ll e^{-\frac{1}{2}\eta_1^2}, \qquad \eta_1 \gg 1, \tag{75}$$

are satisfied. Thus when we have a large number of cells or channels which have very low probability of receiving signal ($d_0 \ll 1$) and where the probability of a noise sample exceeding the peak value of the signal is much greater than the probability of occurrence of signal in a channel ($e^{-\frac{1}{2}\eta_1^2} \gg d_0$) and where the peak signal-to-noise ratio is very large, a threshold detector law is approximately the best. We see that power integration or equivalent methods are not as good as a detector which weights the peaks heavily.

Stated in another way, the problem considered here is equivalent to considering two Bernoulli random variables with threshold probabilities,

$$p_0 = e^{-\frac{1}{2}\eta_1^2}, \tag{76}$$

$$p_1 = d_0 + e^{-\frac{1}{2}\eta_1^2}. \tag{77}$$

We will now prove this. This also serves as a verification of the results which have just been obtained.

For a Bernoulli test the observable which is measured is given by [2], [3], and [8],

$$z(n) = k_n\left[\log \frac{p_1}{p_0} + \log \frac{1-p_0}{1-p_1}\right] - n \log \frac{1-p_0}{1-p_1}. \tag{78}$$

From (76) and (77) and using the Taylor series expansion and approximation to the logarithm we obtain,

$$\log \frac{p_1}{p_0} = \log\,[1 + d_0 e^{\frac{1}{2}\eta_1^2}] \approx d_0 e^{\frac{1}{2}\eta_1^2}, \tag{79}$$

since we have assumed $d_0 \ll e^{-\frac{1}{2}\eta_1^2}$. Similarly,

$$\log \left(\frac{1-p_0}{1-p_1}\right) = -\log\left(1 - \frac{d_0}{1 - e^{-\frac{1}{2}\eta_1^2}}\right) \approx d_0 \tag{80}$$

since,

$$d_0 \ll e^{-\frac{1}{2}\eta_1^2} \ll 1.$$

From (78) and (80) we obtain,

$$z(n) = k_n[d_0 e^{\frac{1}{2}\eta_1^2} + d_0] - n\, d_0,$$
$$\approx k_n\, d_0 e^{\frac{1}{2}\eta_1^2} - n\, d_0, \tag{81}$$

as $e^{\frac{1}{2}\eta_1^2} \gg 1$. Thus (81) derived for the Bernoulli case is exactly the same as (72) which was previously developed under the same set of approximations. We can now state that under the conditions previously stated the best detector given by sequential theory for detecting a sine-wave carrier in noise of large peak-signal-to-noise ratio and very small duty ratio is equivalent to the detector law corresponding to the detection of a random sequence of ones of probability d_0 interleaved (with no overlap) with a random sequence of ones of probability $e^{-\frac{1}{2}\eta_1^2}$ such that $d_0\, e^{\frac{1}{2}\eta_1^2} \ll 1$. The random sequence of ones of probability $e^{-\frac{1}{2}\eta_1^2}$ is considered as the noise sequence while the interleaved sequence of probability d_0 is the signal

sequence. Since $d_0 \ll e^{-\frac{1}{2}\eta_1^2}$ there are many more noise pulses than the desired signal pulses. Thus we actually have another example of a signal embedded in impulse noise with an equivalent signal-to-noise ratio which can be defined by the ratio of the signal and noise probabilities. The equivalent signal-to-noise ratio is $d_0\, e^{\frac{1}{2}\eta_1^2} \ll 1$. The inequality $d_0\, e^{\frac{1}{2}\eta_1^2} \ll 1$ can be written as $\frac{1}{2}\eta_1^2 \ll -\log d_0$. Hence, when the information gained by observing a pulse is much greater than the peak-signal-power-to-noise-power ratio of the pulse a threshold detector is a good approximation to the best detector. The measurement process then involves only counting. This holds when $\eta_1 \gg 1$. The slicing threhold can be set at approximately $r_0 = \eta_1 - 3\sqrt{2}$. This implies that when a pulse is present it will exceed r_0 almost certainly. That is,

$$p_{\eta_1}\{r \geq \eta_1 - 3\sqrt{2}\} \geq 0.90; \qquad \eta_1 \gg 1.$$

Bibliography

[1] Blackwell, D., and Girshick, M. A. *Theory of Games and Statistical Decisions*, New York: John Wiley and Sons, Inc., 1954.

[2] Blasbalg, H. "Theory of Sequential Filtering and Its Application to Signal Detection and Classification," Baltimore: The Johns Hopkins University, Radiation Laboratory, Technical Report No. AF-8, October, 1954.

[3] Blasbalg, H. *Theory of Sequential Filtering and Its Application to Signal Detection*, Doctor of Engineering Dissertation, The Johns Hopkins University, Baltimore, Md. 1956.

[4] Bussgang, J. J., and Middleton, D. "Optimum Sequential Detection of Signals in Noise," IRE Transactions, Vol. IT-1 (December, 1955), pp. 5-18.

[5] Goldman, S. *Frequency Analysis Modulation and Noise*, New York: McGraw-Hill Book Co., Inc., pp. 336-338, 1948.

[6] Middleton, D. "Statistical Criteria for the Detection of Pulsed Carriers in Noise, I, II," *Journal of Applied Physics*, Vol. 24 (April, 1953), pp. 371-391.

[7] Rice, S. O. "Mathematical Analysis of Random Noise," *Bell System Technical Journal*, Vol. 24, No. 1, 1945.

[8] Wald, A. *Sequential Analysis*, New York: John Wiley and Sons, Inc., 1947.

[9] *Note*: A detailed bibliography is given in [4.]

Corrections

Robert Price, author of "Optimum Detection of Random Signals in Noise, with Application to Scatter-Multipath Communication, I," which appeared on pages 125–135 of the December, 1956, issue of these Transactions, has requested the following corrections.

In footnote 9, section 1, the fifth line should read "if the detector output consists of $\cdots$"

In Fig. 1 on page 127, change M to N in the scatter paths.

On page 129, in the first paragraph following (26), the second sentence should begin, "The simultaneous integral equation (22) $\cdots$"

In footnote 23 on page 130, the book by Courant and Hilbert should be identified as Volume I.

On page 131 the term in parentheses on the last line of (54) should be $\phi_1(\sigma - t)$.

In (60) on page 132, change the italic capital I to script capital $\mathcal{I}$. The same correction should be made in the tabulation in the same column on the third line under the heading "This Paper."

In (66), replace 8 by g in the denominator.

Insert "See (4.35) and (4.36)," at the end of the first paragraph of footnote 26. In the same footnote, the first term of the first two equations should be $F_k^{1/2}$.

In (69) change $=$ on the first line to $\equiv$, and change $r^{(2)}$ on the second line to $r^{(k)}$.

On page 133, the heading on the fourth line of the second column should be identified as Case II.

On page 134, two lines below (75), insert "η is given by (65). As $\mu \rightarrow 0$," etc., and on the following line delete "(next page)" at the end of the line.

Immediately below Fig. 5, the last word on the first line should be "least."

In (76) change the italic capital I to script capital $\mathcal{I}$.

David Middleton, author of "On the Detection of Stochastic Signals in Additive Normal Noise—Part I," which appeared on pages 86-121 of the June, 1957, issue of these Transactions, has requested the following corrections.

In (24) on page 90, $\mathbf{k}_N^{-1}$ should immediately follow $\tilde{\mathbf{v}}$.

In (74) on page 95, replace y by x in the upper limit.

On the same page, seven lines below Fig. 1, replace "present" by "preset."

In footnote 18 on page 96, the third line below the equation should begin: by $K_s(|\,t - \tau\,|)$.

In (98), replace B_0^2 in the integrand by B_0.

On page 100, the term $\log(\mathcal{K}/\Gamma_0)$ which appears in (101a), (101b), (103a), and (103b) should be $\log \mathcal{K} - \Gamma_0'$.

In (121) on page 102, the term $\mathbf{K}_{N_0}$ at the end of the first line should be $\mathbf{K}_{N_0}^{-1}$.

In the first line of (128a) on page 103, the term above $\sum$ should be $n +$; the last symbol in the exponent should be $\lambda'_{k,n}$.

On page 106, three lines below (148), T/n should be enclosed in parentheses.

In (168a) on page 109, the second $\tilde{z}$ should be $\tilde{\mathbf{z}}$.

The last line of (187) on page 111 should begin with a + sign. On the same page, $(-1)^k$ should be inserted before $B^{-k/2-1/2}$ on the second line of (188).

In (222) on page 115, the term before the equal sign hould be $\exp\{-p_n \mid t \mid\}$.

On the same page, replace A by A' in the first equation f (225). Here A' is determined by regarding the second quation of (225) as an identity. It is easily seen that $A' = 2A$. Replace A by A' in the following, unless otherwise indicated.

Also on page 115, on the second line of (229), the symbol ollowing V_T should be $(t - t')$.

In (230), (231), (233), and (236) on page 116, replace A by $A' = 2A$.

On page 116, on the second and third lines of (231a), he last term should be $e^{-c_n t}$. In (231b) the last term on he second and third lines should be $e^{c_n(t-T)}$.

In (239) delete the 2 in the second relation.

In (241) and (242) on page 117, insert a minus sign before F_n. Also insert a minus sign after dx in the first ine of (243).

In (244), replace the 4's by 2's and the 2 by 1.

In (245a) replace 4 by 2 and place the radical in the denominator for h_1. In the last two relations, insert factor 1/2 and put the radical in the denominator.

In (246) delete the first factor 2. Place the first radical in the denominator, and in the second line of the equation $1 + \gamma_0^2$ should be $\sqrt{1 + \gamma_0^2}$.

In (247a) and (247b), multiply by 1/2.

In (248) delete the factor 2, and place the first radical in the denominator, outside the parentheses. The second term should be preceded by a minus sign.

On page 118, in (254) delete the factors 2 and place an equal sign between h_1 and h_2^* on the second line of the equation.

In (256) replace the factor 8 by 4, insert minus signs before c_1 in the last two exponents, and insert a minus sign before the second term.

On the first line of (268), delete the minus sign before the last exponent $c_n(t - T)$.

In (269a) and (270) the first $(b_n + c_n)^2$ should be $(b_n - c_n)^2$. Insert a minus sign before the second term in (270).

Correspondence

The Sample Space Trajectory of Time-Shifted Signal Vectors*

While investigating optimum phase-determining filters,[1] it was convenient to give some geometric significance to the trajectory traced by the tip of the vector representing a band-limited, time-shifted signal in sample space.

One constraint on the tip trajectory of such a signal vector has been stated by Shannon.[2] This requires the signal to remain on the surface of a hyper-sphere of radius

$$r = \sqrt{2WE} \tag{1}$$

where

W = highest frequency in the signal
E = total energy in the signal.

It may be demonstrated that another constraint on the trajectory of a time-shifted signal is that it must lie on a hyper-plane having identical direction cosines (or having equal intercepts on all coordinate axes). The perpendicular distance between the hyper-plane containing the time-shifted signal and the origin is proportional to the integrated area under the signal.

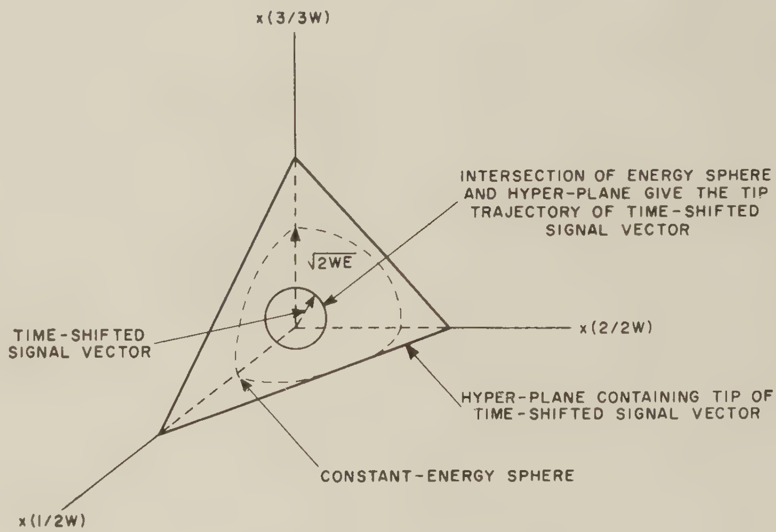


Fig. 1.

Note that[2] for a band-limited signal $x(t)$ translated by τ

$$x(t - \tau) = \sum_{i=-\infty}^{\infty} x\left(\frac{i}{2W} - \tau\right) \cdot \frac{\sin W(2\pi t - i)}{W(2\pi t - i)}.$$

Integrating and using the known value of the sine-integral gives

$$\sum_{i=-\infty}^{+\infty} x\left(\frac{i}{2W} - \tau\right) = 2W \int_{-\infty}^{+\infty} x(t)\, dt. \tag{2}$$

Eq. (2) represents the equation of the hyper-plane to which the tip of the time-shifted signal vector is constrained. This hyper-plane has equal direction cosines (or equal intercept values with all coordinate axes). The distance, along a normal, from the origin to the hyper-plane, which is set by the expression on the right-hand side of (2), is proportional to the integrated area under the signal.

One can view the intersection either as a

* Received by the PGIT, May 23, 1957.

[1] This work appears as a portion of ch. 5 of Dissertation entitled, "The Detection and Phase Determination of Signals in Additive and Multiplicative Noise," submitted June, 1955, to Polytechnic Institute of Brooklyn, Brooklyn, N. Y., in partial fulfillment of the requirements for the author's D.E.E. degree.

[2] C. E. Shannon, "Communication in the presence of noise," Proc. IRE, vol. 37, pp. 10-21; January, 1949.

trajectory along the surface of the constant energy hypersphere or as a path described by the signal point in hyper-space as the axes are rotated.

The intersection of the hyper-plane and hyper-sphere is a hyper-sphere reduced in dimensionality[3] by 1. (See the three-dimensional periodic signal example in Fig. 1, p. 257.) This has delimited the trajectory only very slightly, but it is useful in confirming our beliefs concerning those elements of a signal which *do not* carry phase information. As the integrated area under the signal becomes larger, the hyper-plane is just tangent to the hyper-sphere at one point. This point corresponds to the signal whose time-shift causes all samples to rotate into themselves (*e.g.*, direct current).

[3] In the limit of infinite dimensionality, the resulting infinite energy must be normalized.

In order to make the points on the trajectory as disparate as possible, the integrated area under the signal should be zero. This should give the largest possible radius to the hyper-sphere of reduced dimensionality created by the intersecting surfaces. This confirms our intuitive notions that direct current (or integrated value of the signal) carries no phase information and should, therefore, be avoided in generating limited energy signals whose phase or time-shift is to be measured.

In a limited sample space of three dimensions, it is interesting to note that all time-shifted signals of three dimensions are identical except in magnitude and direct current content.

The remaining parameters needed to describe the tip trajectory are the Fourier coefficients which define additional hyper-planes in sample space and further constrain the intersection with the constant-energy hyper-sphere.

Herbert Sherman
Lincoln Lab.
Mass. Inst. Tech.
Lexington, Mass.

Contributors

Robert I. Bernstein (S'51—A'52—SM'57) was born in New York, N. Y., on July 13, 1927. He received the B.E.E. degree from the City College of New York in 1948, the M.E.E. degree from New York University in 1952, and the D. Sc. Eng. degree from Columbia University in 1955.

R. I. Bernstein

He joined the Columbia University Electronics Research Laboratories in 1952, after being with the Fairchild Guided Missiles Division for several years. He is the Associate Director of the Laboratories, an associate professor of engineering, and has been in charge of the ORDIR long-range coherent radar project since its conception.

Dr. Bernstein is a member of Sigma Xi, Tau Beta Pi, and Eta Kappa Nu.

❖

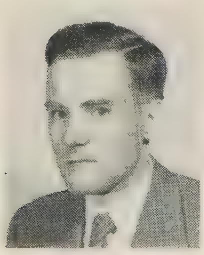
For a photograph and biography of Marvin Blum, see page 208 of the September, 1957, issue of these Transactions.

❖

For a photograph and biography of Herman Blasbalg, see pp. 168-169 of the June, 1957, issue of these Transactions.

❖

Carl W. Helstrom was born in Easton Pa., on February 22, 1925. From 1944 to 1946 he served as a radio technician in the United States Navy. In 1947 he received the B. S. degree in engineering physics from Lehigh University and in 1951 the Ph. D. degree in physics from the California Institute of Technology. Since 1951 he has been employed at the Westinghouse Research Laboratories, Pittsburgh, Pa., where he is now a member of the Department of Mathematics.



C. W. Helstrom

Dr. Helstrom is a member of Phi Beta Kappa and the American Physical Society.

❖

Kenneth S. Miller (A '47—M '52—SM '57) was born on June 4, 1922, in New York, N. Y. He received the B.S. degree in chemical engineering and the A. M. and Ph. D. degrees in mathematics from Columbia University. His postdoctoral work was done at the Institute for Advanced Study, Princeton, N. J., in 1950.

K. S. Miller

During World War II, Dr. Miller served as a radar officer in the U. S. Navy. Since then he has acted as a consultant to various industrial and governmental agencies on problems associated with system analysis, noise, mathematical machines, and applied mathematics. He is the author or co-author of numerous research papers and seven books. At present, he is an associate professor of mathematics at New York University.

Dr. Miller is a member of the American Mathematical Society, Sigma Xi, Tau Beta Pi, and Pi Mu Epsilon.

❖

Marvin Shinbrot was born in Brooklyn, N. Y., in 1928. He received the Bachelor's degree in mathematics in 1948 and the Master's degree in the same subject in 1949 from Syracuse University.

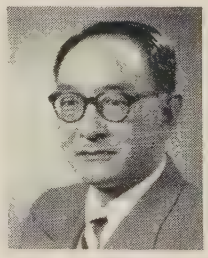
Since November, 1949, Mr. Shinbrot has worked for the National Advisory Committee for Aeronautics as an aeronautical research scientist. His work for the past few years has, to a large extent, been concerned with optimization of systems when the inputs are not stationary.

M. Shinbrot

Mr. Shinbrot is a member of the American Mathematical Society, the Mathematical Association of America, the Society for Industrial and Applied Mathematics, Sigma Xi, and the American Association for the Advancement of Science.

❖

Satosi Watanabe was born in Tokyo, Japan, on May 26, 1910. In 1933, he graduated from the University of Tokyo with the B. Sc. degree. Then he attended the University of Paris, receiving from the French government the D. Sc. degree in 1935. He received the D. Sc. degree in 1940 from the University of Tokyo. From 1941 to 1945, he was an associate professor at the University of Tokyo. During 1949-1950, he held the position of professor and chairman at St. Paul's University, also in Tokyo. In 1950, Dr. Watanabe came to the United States and was employed at Wayne University in Detroit, Mich., as an associate professor, a position he held until 1952. In 1952, he became a professor of the U. S. Naval Postgraduate School in Monterey, Calif. Since 1956 he has been employed as senior physicist at the IBM Research Laboratory in Poughkeepsie and Ossining, N. Y.



S. Watanabe

Index to

IRE TRANSACTIONS

ON

INFORMATION THEORY

Volume IT-3, 1957

IRE Transactions on Information Theory

Index to Volume IT-3—1957

Contents

Volume IT-3, Number 1, March, 1957

Index Number		Page
	Frontispiece, *P. M. Woodward*	2
IT143.	Entropy and Negentropy (Editorial), *P. M. Woodward*	3
	The IRE Affiliate Plan—A New Venture in Engineering Society Structure and Service, *W. R. G. Baker*	4
IT144.	On the Estimation in the Presence of Noise of the Impulse Response of a Random Linear Filter, *G. L. Turin*	5
IT145.	The Output Signal-to-Noise Ratio of Correlation Detectors, *P. E. Green, Jr.*	10
IT146.	Error Rates in Pulse Position Coding, *L. L. Campbell*	18
IT147.	The Part of Statistical Considerations in the Separation of a Signal Masked by a Noise, *J. A. Ville*	24
IT148.	On a Cross-Correlation Property for Stationary Random Processes, *J. L. Brown, Jr.*	28
IT149.	A Systematic Approach to a Class of Problems in the Theory of Noise and Other Random Phenomena—Part I, *D. A. Darling and A. J. F. Siegert*	32
IT150.	A Systematic Approach to a Class of Problems in the Theory of Noise and Other Random Phenomena—Part II, Examples, *A. J. F. Siegert*	38
IT151.	On the Capacity of a Noisy Continuous Channel, *S. Muroga*	44
IT152.	Merit Criteria for Communication Systems, *A. Hauptschein and L. S. Schwartz*	52
IT153.	First-Order Markov Process Representation of Binary Radar Data Sequences, *G. C. Sponsler*	56
IT154.	Automatic Bias Control for a Threshold Detector, *J. Dugundji and E. Ackerlind*	65
IT155.	Exact Integral Equation Solutions and Synthesis for a Large Class of Optimum Time Variable Linear Filters, *J. S. Bendat*	71
	Contributors	81

Volume IT-3, Number 2, June, 1957

Index Number		Page
	Frontispiece, *F. L. H. M. Stumpers*	84
IT156.	Information Theory and International Radio Organizations (Editorial), *F. L. H. M. Stumpers*	85
IT157.	On the Detection of Stochastic Signals in Additive Normal Noise—Part I, *D. Middleton*	86
IT158.	The Relationship of Sequential Filter Theory to Information Theory and Its Application to the Detection of Signals in Noise by Bernoulli Trials, *H. Blasbalg*	122
IT159.	A Theory of Weighted Smoothing, *L. A. Ule*	131
IT160.	The Response of Phase-Locked Loop to a Sinusoid Plus Noise, *S. G. Margolis*	136
IT161.	A Note on the Sampling Principle for Continuous Signals, *A. V. Balakrishnan*	143
IT162.	A Note on Some Statistics Concerning Typewritten or Printed Material, *S. Deutsch*	147
IT163.	A Note on the Construction of a Multivariate Normal Sample, *G. Marsaglia*	149
IT164.	A Bibliography of Information Theory (Communication Theory—Cybernetics) (Second Supplement), *F. L. H. M. Stumpers*	150
	Correspondence:	
IT165	Determination of Redundancies in a Set of Patterns, *E. C. Riekeman, A. Glovazky, and E. J. McCluskey, Jr.*	167
	Contributors	168

Volume IT-3, Number 3, September, 1957

Index Number		Page
	Frontispiece, *B. McMillan*	172
IT166.	Where Do We Stand? (Editorial), *B. McMillan*	173
IT167.	Detection of Fluctuating Pulsed Signals in the Presence of Noise, *P. Swerling*	175
IT168.	Fixed Memory Least Squares Filters Using Recursion Methods, *M. Blum*	178
IT169.	Locally Stationary Random Processes, *R. A. Silverman*	182
IT170.	The Solution of a Homogeneous Wiener-Hopf Integral Equation Occurring in the Expansion of Second-Order Stationary Random Functions, *D. C. Youla*	187
IT171.	The Correlation Function of Smoothly Limited Gaussian Noise, *R. F. Baum*	193
IT172.	On the Role of Dynamic Programming in Statistical Communication Theory, *R. Bellman and R. Kalaba*	197
IT173.	Complex Processes for Envelopes of Normal Noise, *R. Arens*	204
	Correspondence:	
IT174.	A Question of Terminology, *J. G. Kreer*	208
	Contributors	208

Volume IT-3, Number 4, December, 1957

Index Number		Page
	Frontispiece, *W. B. Davenport, Jr.*	212
IT175.	Sputnik Et Cetera (Editorial), *W. B. Davenport, Jr.*	213
IT176.	A Theory of Multilevel Information Channel with Gaussian Noise, *S. Watanabe*	214
IT177.	A Generalization of a Method for the Solution of the Integral Equation Arising in Optimization of Time-Varying Linear Systems with Nonstationary Inputs, *M. Shinbrot*	220
IT178.	On the Mean Square Noise Power of an Optimum Linear Discrete Filter Operating on Polynomial Plus White Noise Input, *M. Blum*	225
IT179.	The Distribution of the Number of Crossings of a Gaussian Stochastic Process, *C. W. Helstrom*	232
IT180.	An Analysis of Coherent Integration and Its Application to Signal Detection, *K. S. Miller and R. I. Bernstein*	237
IT181.	The Sequential Detection of a Sine-Wave Carrier of Arbitrary Duty Ratio in Gaussian Noise, *H. Blasbalg*	248
IT182.	Correction to "Optimum Detection of Random Signals in Noise, with Application to Scatter-Multipath Communication, I," *R. Price*	256
IT183.	Correction to "On the Detection of Stochastic Signals in Additive Normal Noise—Part I," *D. Middleton*	256
	Correspondence:	
IT184.	The Sample Space Trajectory of Time-Shifted Signal Vectors, *H. Sherman*	257
	Contributors	258
	Annual Index 1957 . . . . Follows Page	258

Index to Authors

Numbers refer to index numbers in Contents listing

A

Ackerlind, E.: IT154
Arens, R.: IT173

B

Balakrishnan, A. V.: IT161
Baum, R. F.: IT171
Bellman, R.: IT172
Bendat, J. S.: IT155
Bernstein, R. I.: IT180
Blasbalg, H.: IT158, IT181
Blum, M.: IT168, IT178
Brown, J. L., Jr.: IT148

C

Campbell, L. L.: IT146

D

Darling, D. A.: IT149
Davenport, W. B., Jr.: IT175
Deutsch, S.: IT162
Dugundi, J.: IT154

G

Glovazky, A.: IT165
Green, P. E., Jr.: IT145

H

Hauptschein, A.: IT152
Helstrom, C. W.: IT179

K

Kalaba, R.: IT172
Kreer, J. G.: IT174

M

Margolis, S. G.: IT160
Marsaglia, G.: IT163
McCluskey, E. J., Jr.: IT165
McMillan, B.: IT166
Middleton, D.: IT157, IT183
Miller, K. S.: IT180
Muroga, S.: IT151

P

Price, R.: IT182

R

Riekeman, E. C.: IT165

S

Schwartz, L. S.: IT152
Sherman, H.: IT184
Shinbrot, M.: IT177
Siegert, A. J. F.: IT149, IT150
Silverman, R. A.: IT169
Sponsler, G. C.: IT153
Stumpers, F. L. H. M.: IT156, IT164
Swerling, P.: IT167

T

Turin, G. L.: IT144

U

Ule, L. A.: IT159

V

Ville, J. A.: IT147

W

Watanabe, S.: IT176
Woodward, P. M.: IT143

Y

Youla, D. C.: IT170

Index to Technical Subjects

B

Bernoulli Detection: IT159
Bibliographies, on Information Theory: IT164

C

Capacity of Noise Continuous Channel: IT151
Carrier, Sine-Wave, Sequential Detection of: IT181
Channels: IT151, IT152, IT176
  Multilevel Information, with Noise: IT176
  Noisy: IT151, IT152
    Continuous Capacity: IT151
    Merit Criteria for Information Transmission: IT152
Coding, Pulse Position, Error Rates: IT146
Coherent Integration Applied to Detection: IT180
Co-Information, Question of Terminology: IT174
Communications Systems, Merit Criteria: IT152
Communication Theory: IT164, IT172
  Bibliography on: IT164
  Statistical Dynamic Programming in: IT172
Continuous Signals, Sampling Principle for: IT161
Correlation Detectors, Output Signal-to-Noise Ratio: IT145
Correlation Function of Smoothly Limited Gaussian Noise: IT171
Cross-Correlation Property for Stationary Random Processes: IT148
Crossings of a Gaussian Stochastic Process, Distribution of: IT179
Cybernetics, Bibliography on: IT164

D

Data Smoothing, Weighted: IT159
Detection: IT147, IT157, IT158, IT167, IT180-IT183
  Bernoulli, and Sequential Filter Theory: IT158
  Coherent Integration Applied to: IT180
  Gaussian Signals in Noise for Scatter-Path Communication, Correction to IT137: IT182
  Pulsed Signal in Presence of Noise: IT167
  Sequential, of a Sine-Wave Carrier: IT181
  Signal Masked by Noise: IT147
  Stochastic Signals in Additive Normal Noise: IT157, IT183
    Correction: IT183
Detectors: IT145, IT154
  Automatic Control of Threshold Bias: IT154
  Correlation, Output Signal-to-Noise Ratio: IT145

E

Entropy, Negentropy: IT143
Envelopes of Normal Noise, Complex Processes for: IT173
Error Rates in Pulse Position Coding: IT146

F

Facsimile, Probability Statistics Concerning Typewritten or Printed Material: IT162
Filters: IT144, IT155, IT158, IT159, IT168, IT178
  Fixed Memory Least Squares, Using Recursion Methods: IT168
  Linear: IT144, IT155
    Optimum Time Variable, Synthesis of: IT155
    Random, Impulse Response Presence of Noise: IT144
  Optimum Linear, Noise Power of: IT178
  Sequential, Signal Detection by Bernoulli Trials: IT158
  Weighted Smoothing: IT159

G

Gaussian Noise, Smoothly Limited, Correlation Function of: IT171
Gaussian Sample, Generating of, Construction of Multivariate Normal Sample: IT163
Gaussian Stochastic Process, Distribution of Crossings: IT179
Generating a Gaussian Sample, Construction of Multivariate Normal Sample: IT163

I

Information, Multilevel Channel, with Noise: IT176

Information Theory: IT156, IT164, IT166
Bibliography on: IT164
International Radio Organizations: IT156
Where Do We Stand? IT166
Information Transfer, Question of Terminology: IT174
Integration, Coherent, Applied to Detection: IT180
International Radio Organizations, and Information Theory: IT156

L

Least Squares Filters, Fixed Memory, Using Recursion Methods: IT168
Linear Systems, Time-Varying, Optimization of: IT177

M

Markov Process Representation of Radar Data: IT153
Merit Criteria for Noisy Channels: IT152
Multilevel Information Channel with Noise: IT176
Mutual Information, Question of Terminology: IT174

N

Negentropy: IT143
Noise: IT144, IT147, IT149, IT150, IT157, IT160, IT171, IT173, IT176, IT178, IT181-IT183
Complex Processes for Envelopes of: IT173
Detection: IT157, IT182, IT183
of Band-Pass Signals for Scatter-Path Communication, Correction to IT137: IT182
of Stochastic Signals: IT157, IT183
Correction: IT183
Gaussian, Smoothly Limited, Correlation Function of: IT171
Multilevel Information Channel with: IT176
Phase-Locking in Presence of: IT160
Power of an Optimum Linear Discrete Filter: IT178
Random Linear Filter, Estimation of Impulse Response: IT144
Sequential Detection of a Sine-Wave Carrier: IT181
Statistical Considerations Involved in Separation of Signal: IT147
Systematic Approach to Class of Problems: IT149, IT150
Examples: IT150

O

Optimization of Time-Varying Linear Systems: IT177
Oscillators, Voltage-Controlled, Phase Locking: IT160

P

Pattern Redundancies, Determination of: IT165
Phase-Locked Loop, Response to Sinusoid Plus Noise: IT160
Printed Material, Scanning, Statistics Concerning: IT162
Programming, Dynamic, in Statistical Communication Theory: IT172
Pulsed Signals, Detection in Presence of Noise: IT167
Pulse Position Coding, Error Rates: IT146

R

Radar, Markov Process Representation of Data: IT153
Radio Organizations, International, and Information Theory: IT156
Random Processes: IT169, IT170
Locally Stationary: IT169
Wiener-Hopf Integral: IT170
Redundance in Set of Patterns, Determination of: IT165

S

Sample Space Trajectory of Time-Shifted Signal Vectors: IT184
Sampling: IT161, IT163
Multivariate Normal: IT163
Theorems for Continuous Signals: IT161
Scanning: IT162, IT165
Pattern Redundancies, Determination of: IT165
Probability Statistics Concerning Typewritten or Printed Material: IT162
Sequential Filter Theory, Bernoulli Detection: IT158
Solution of Linear System Optimization Equations: IT177
Stationary Random Processes: IT148, IT170
Cross-Correlation Property: IT148
Wiener-Hopf Integral: IT170
Statistical Communication Theory, Dynamic Programming in: IT172
Statistics Concerning Typewritten or Printed Material: IT162
Stochastic Process: IT157, IT179, IT183
Gaussian, Distribution of Crossings: IT179
Signal Detection in Additive Normal Noise: IT157, IT183
Correction: IT183

T

Threshold Detector, Automatic Bias Control: IT154
Time-Varying Linear Systems, Optimization of: IT177
Trajectory, Sample Space, of Time-Shifted Signal Vectors, IT184
Transformation, Question of Terminology: IT174
Typewritten Material, Scanning, Probability Statistics Concerning: IT162

V

Vectors, Time-Shifted Signal, Space Trajectory of: IT184

W

Weighted Smoothing, Theory of: IT159
Where Do We Stand?: IT166
Wiener-Hopf Integral in Stationary Random Functions: IT170

INFORMATION FOR AUTHORS

Authors are requested to submit editorial correspondence or technical manuscripts to the Publications Chairman for possible publication in the PGIT Transactions. Papers submitted should include a statement as to whether the material has been copyrighted, previously published, or accepted for publication elsewhere.

Papers should be written concisely, keeping to a minimum all introductory and historical material. It is seldom necessary to reproduce in their entirety previously published derivations, where a statement of results, with adequate references, will suffice.

To expedite reviewing procedures, it is requested that authors submit the original and two legible copies of all written and illustrative material. The manuscript should be double-spaced, and the illustrations drawn in India ink on drawing paper or drafting cloth. Each paper should include a carefully written abstract of not more than 200 words. Upon acceptance, papers should be prepared for publication in a manner similar to those intended for the Proceedings of the IRE. Further instructions may be obtained from the Publications Chairman. Material not accepted for publication will be returned.

IRE Transactions on Information Theory is published four times a year, in March, June, September, and December. A minimum of one month must be allowed for review and correction of all accepted manuscripts. In addition, a period of approximately two months is required for the mechanical phases of publication and printing. Therefore, all manuscripts must be submitted three months prior to the respective publication dates. In addition, the IRE National Convention Record is published in July, and the IRE WESCON Convention Record in the Fall. A bound collection of Information Theory papers delivered at these conventions is mailed gratis to all PGIT members.

All technical manuscripts and editorial correspondence should be addressed to Laurin G. Fischer or George A. Deschamps, Federal Telecommunication Labs., 492 River Road, Nutley 10, N. J. Local Chapter activities and announcements, as well as other nontechnical news items, should be addressed to Nathan Marchand, Marchand Electronic Labs., Riversville Road, Greenwich, Conn.